Alien plants and their invasion of the forested landscape of the southeastern United States

A thesis submitted in
partial fulfilment
of the requirements for
the Degree of
Doctor of Philosophy in Environmental Science
at
Canterbury University

by
Dawn Lemke

Supervisors:
Jennifer Brown
Philip Hulme

Table of Contents

List of Figures

List of Tables

Acronyms

AIC - Akaike information criterion

ASMC – Alabama Surface Mining Commission

AUC – Area Under the Curve

CCA – Canonical Correspondence Analysis

CIR – Colour Infrared

CPMR – Cumberland Plateau and Mountain Region

DBH – Diameter at Breast Height

DEM – Digital Elevation Model

FIA – Forest Inventory and Analysis

FN – False Negative

GBIF - Global Biodiversity Information Facility

GCM – Global Circulation Model

GRTS - Generalized Random Tessellation Stratified

LULC – Land Use Land Cover

NDVI – Normalized Difference Vegetation Index

NLCD – National Land Cover Database

ROC – Receiver Operator Characteristic

SD – Standard Deviation

SHR – Shale Hills Region

SMCRA - Surface Mining Control and Reclamation Act

SDM – Species Distribution Models

TN – True Negative

TP – True Positive

USDA – United Stated Department of Agriculture

USFS – United Stated Forest Service

USGS– United Stated Geological Service

Acknowledgments

First and foremost, I would like to thank my supervisors Jennifer Brown and Philip Hulme, for giving their valuable time, encouraging me through the challenges, showing me balance is possible and for giving me this opportunity. I am grateful to my collaborators in this process, Wubishet Tadesse, Yong Wang, Callie Schweitzer, John Coulston, Jim Miller, and Irenus Tazisong, all of whom have been great to interact with and to learn from. I would like to thank Dana Virone, Shelley Baltar, Matthew Zirbel, Clinton Patterson, Breana Cobb, Ryan Sisk and Matt Carr for their assistance in the field and in the lab. I would like to express my appreciation to my examiners for spending time on my thesis and for their valuable comments. I would also like to thank University of Canterbury, United States National Science Foundation, United States Office of Surface Mining and the United States Department of Agriculture Forest Service, whose funding allowed this to come together.

Many thanks to the students, staff and faculty, at Alabama Agricultural and Mechanical University and University of Canterbury, for their assistance and for making my time during this process so enjoyable, especially Kathy Roberts and Lisa Gardner.

Finally, I extend my deepest appreciation to my family and friends, who have graciously coped with all of my absences and distractions, my angst and frustrations, and have helped and encouraged me throughout my studies. I am especially grateful to Scott, Drew and Ian.

Thesis Structure

Each chapter was written in a manuscript style, thus there is some repetition in methods, particularly in species, study area descriptions and modelling techniques. Collaborators outside of my supervisors are acknowledged at the beginning of each chapter (Table 1). The research chapters are prefaced by a short introduction chapter and I conclude with a final summary chapter. The majority of the work is my own, completed with guidance from collaborators and supervisors. The soils lab analysis from Chapter 2 was undertaken by Dr Tazisong, who graciously allowed me to integrate the results into the chapter.

Table 1: Detailed list of contributors to Dawn Lemke PhD research. All work was primarily Ms. Lemke's but was developed with discussion with many mentors. All soils analysis was conducted by Dr. Tazisong.

Person	Affiliation	Contribution
Jennifer Brown	Canterbury	As major supervisor, contributed to all chapters, in development of concepts and revision of manuscripts.
Philip Hulme	Lincoln	As supervisor, contributed primarily to Chapters 1 and 4 through 8, in development of concepts and revision of manuscripts.
Yong Wang	Alabama A&M	Assisted in obtaining funding through the U.S. Office of Surface Mining to undertake work in Chapters 2 and 3. Assisted in development of overall concepts and revision of manuscripts for Chapters 2 and 3.
Callie Schweitzer	US Forest Service	Assisted in obtaining funding through the U.S. Office of Surface Mining to undertake work in Chapters 2 and 3. Assisted in development of overall concepts and revision of manuscripts for Chapters 2 and 3, specifically in a forestry context. Assisted with the initial field methods.
Wubishet Tadesse	Alabama A&M	Assisted in development of overall concepts and revision of manuscripts for Chapters 2, 3 and 4, specifically in a remote sensing context.
Irenus Tazisong	Alabama A&M	Assisted in development of overall concepts and revision of manuscripts for Chapters 2 and 3, specifically in a soils context. Completed all soils analysis in Chapter 2.
John Coulston	US Forest Service	Assisted in development of overall concepts and revision of manuscripts for Chapters 6 and 7. Assisted in obtaining funding from the USFS for work undertaken in Chapters 6 and 7 and the USFS Forest Futures Report. Extracted relevant data from the FIA database.
Jim Miller	US Forest Service	Assisted in development of overall concepts and revision of manuscripts for Chapters 6 and 7, specifically focused on the application to invasive plant management. Assisted in obtaining funding from the USFS for work undertaken in Chapters 6 and 7 and the USFS Forest Futures Report.

Abstract

In this thesis, I have assessed and modelled invasion of alien plant species in the forest of the southeastern United States. There are over 380 recognized invasive plants in southeastern forests and grasslands with 53 ranked as high-to-medium risk to natural communities. I have focused on ten of these: Chinese lespedeza, tall fescue, Japanese honeysuckle, Chinese privet, autumn olive, princesstree, silktree, chinaberry, tree of heaven, tallowtree. Assessing them at differing scales, locally (Chapter 2 and 3), eco-regionally (Chapter 4 and 5) and regionally (Chapters 6 and 7), using field based measurements integrated with remotely sensed and digital datasets, and applying both parametric and non-parametric modelling approaches. Data from field based measurements as well as digitally available sources was evaluated, bringing together freely available data with time consuming, intensively collected data. Once models were developed application to assessing long term impacts was done by integrating potential climate change scenarios.

At the local level Chinese lespedeza and Japanese Honeysuckle were the most prevalent, with models at the local level dominated by remotely sensed variables. At an eco-regional level Japanese honeysuckle was the most prevalent with models primarily dominated by environmental variables. At a regional level, where only trees were assessed, potential distributions of the invasive species ranged from 12 to 33 percent of the southeastern forests under current conditions with this dramatically increasing for chinaberry and tallowtree under most climate change scenarios, up as high as 66 percent of southeastern forest sites.

In this thesis information on anthropogenic factors added some value to the models, however it was rarely dominant. Roads and land use (proportion of forest or distance to forest) were the most useful anthropogenic variables. In all models evaluated, only six times did any one anthropogenic variable represent more than 25 percent of the models, four of these were at the local scale. At the regional and eco-regional level, roads had a greater than 25 percent contribution to the silktree models, at a local level, distance to forest and distance roads contributed more than 25 percent to three of the species evaluated, sawtooth oak, Japanese honeysuckle and privet. Human activities have the most influence on invasion progression through dispersal (movement and introduction

rate) and disturbance of the landscape (increased resource availability). Anthropogenic variables such as roads are likely to be a mechanism of spread, thus the more a model is driven by anthropogenic variables, the more likely the invasive plant is to be in the early stages of invasion process. Thus our results suggest that many of these species have moved through the first stages of invasion.

Environmental characteristics play an important role in determining a site's vulnerability to invasion. At an eco-region and regional scale, environmental characteristics dominated (>50%) all but one model (silktree at the regional scale). At the eco-region level elevation was the dominant variable, and at a regional level minimum temperature was the dominant variable. These have some correlation, with higher elevation often relating to lower temperatures, particularly at a smaller scale. This confirms the validity of matching the climate ranges of native species with the range of potential invasion, and the approach of integrating elevation, latitude and longitude to estimate potential distribution. It also suggests that climate change will influence the distribution and that variation in climate should be integrated into models.

Two different modelling approaches, logistic regression and maximum entropy, were used throughout my thesis, and applied to the same data. Agreement between different modelling types adds strength to conclusions, while disagreement can assist in asking further questions. The inclusion in the models of similar variables with the same direction of relationships gives confidence to any inference about the importance of these variables. The geographical agreement between models adds confidence to the probability of occurrence in the area. Alternatively using the same model but different datasets can give you similar information. Overall for all models created by both logistic regression and MaxEnt, the logistic regression had slightly better omission rates and the MaxEnt model had better AUC's. Logistic regression models also often predicted larger geographical areas of occurrences when the threshold of maximum sensitivity plus specificity was used, thus the lower omission rates is related to the less stringent model that predicts a larger area. The selection of appropriate data to answer the question was shown to be fundamental in Chapter 7. When data were used outside of the area of interest it generalized the models and increased the potential for invasion significantly. There was more value in the intensive surveyed data but this was less dramatic than in using the defined areas of interest to select the data for models.

CHAPTER 1.

Introduction

In this thesis I assess and model the invasion probability of alien plant species in the forest of the southeastern United States. There are over 380 recognized invasive plants in southeastern forests and grasslands (more than 330 terrestrials and 48 aquatics), with 53 are ranked as high-to-medium risk to natural communities (Miller et al. 2010). This work focuses on ten of these and includes trees, shrubs, vines, forbs and grasses.

Invasions are not new phenomena; they have been occurring throughout time (Brown & Sax 2004). However, intentionally and unintentionally, humans have moved thousands of species outside of their native ranges, with massive acceleration in this introduction rate over the last century (Hulme et al. 2009). While human induced invasions are considered by some to be primarily the same as natural invasions (Brown & Sax 2004), human induced invasions differ in that they now occur at a much faster rate and throughout the world, with all continents and oceans being affected simultaneously (Ricciardi 2007). Human induced invasions also differ in that they bring species from well beyond the geographical area. Natural invasions come from neighbouring areas over long periods of time, often occurring in waves after geographic barriers have been lifted. These events are substantially different from human induced invasions in spatial and temporal scales and in the diversity of organisms involved in long-distance dispersal (Ricciardi 2007). The growing human population, rise in global trade, relative ease of travel and transport, and degradation of native habitats have all contributed to the increase in introduction rate and the resulting establishment and spread of alien species (Hulme et al. 2009; Ricciardi 2007). Although most individuals die soon after their release, some of these species become invasive, spreading away from their initial site of establishment (Richardson et al. 2000) often with detrimental effects (ecological and economic) on native and managed ecosystems.

In the United States, alien species cause environmental damage and economic loss in excess of US$120 billion a year, $34 billion per year for plants alone (Pimentel et al. 2005). In 1998 approximately 400 of the 958 species listed as threatened or endangered under the United States Endangered Species Act were at risk due to competition or predation by invasive species (Wilcove et al. 1998). Invasive species are now recognized as a major component of global environmental change (Ricciardi 2007; Vitousek et al. 1997) and one of the greatest threats to

biodiversity second only to habitat destruction (Wilcove et al. 1998). Controlling invasive plants is costly and affects both the viability of the forest industry as a whole as well as ecosystem sustainability (Pimentel et al. 2005).

The impact invasive plants can have on our environment includes replacement of native communities and degradation of ecosystems (Pyšek & Richardson 2010). An invasive plant can be defined as an introduced plant that produces reproductive off-spring in areas distant from sites of introduction, in either disturbed or natural systems (Richardson et al. 2000). Within forest systems they pose one of the most immediate threats to present and future forests (Miller et al. 2010; Moser et al. 2009). The replacement of diverse native plant communities by dense infestations, with limited species diversity, is becoming widespread, altering forest ecosystem structure and function (Holmes et al. 2009; Pfeiffer & Voeks 2008). These infestations can reduce biodiversity, limit wildlife habitat, decrease forest productivity and threaten forest health (Pimentel et al. 2005; Reinhart et al. 2005; Standish et al. 2001; Webster et al. 2006). Alterations to forest structure and natural succession can result in changes in functions and processes of the ecosystem, reducing ecosystem services, including soil formation, water balance, and air purification (Hawkes et al. 2005; Martin et al. 2009; Vitousek et al. 1997). Furthermore, humans have also created assemblages that would never have come about through natural processes and thus creating new evolutionary pressures (Strauss et al. 2006).

There are two broad phases of invasion, introduction, which includes transport and colonization, and persistence, which includes establishment and landscape spread. Introduction requires a source population and opportunity, while persistence requires an ecological space (Theoharides & Dukes 2007). This work focuses on the persistence phase.

The ability of a non-native plant to invade depends not only on the attributes of the plant (innate characteristics), but also on the characteristics of the habitat being invaded (habitat characteristics) (Brown & Peet 2003; Pysek & Richardson 2007). Innate characteristics that often lead to successful invasion include rapid growth, short life cycle, toleration of a wide range of conditions, deep root system, prolific flowering, production of many seeds, long seed dormancy and staggered germination, efficient seed dispersal, ability to reproduce sexually and asexually, and allelopathic characteristics. These are all traits that can assist invasion and the rate at which this invasion occurs. Habitat factors can be broken down to anthropogenic and environmental. Human activities have the most influence on invasion progression through dispersal (movement and introduction rate) and disturbance of the landscape. One of the main vectors for the spread

can be human activities, both intentional (e.g. planting) and unintentional (e.g. accidental movement on equipment) (Hodkinson & Thompson 1997). Thus early in the invasion the species can be seen in higher concentration near roads and around urban areas. Once alien plants are in an area it is often human induced disturbance that gives them the opportunity to flourish. Human activities have significantly altered landscapes through the fragmentation of natural habitats with the development of agricultural and urban land (Fischer & Lindenmayer 2007). This environmental heterogeneity influences whether ecosystems can resist alien species invasions and the rate at which an invasion process will likely occur across the landscape (Melbourne et al. 2007). Although disturbance is not required for invasion, it has been identified as one of the most reliable indicators of vulnerability to plant invasion (Buckley et al. 2007; Pysek 2002). The association between invasion and disturbance has been related to changes in resource availability resulting from the disturbance; the greater the availability of resources the higher chance for establishment (Davis et al. 2000). Historical land use and disturbances can also have a major impact on resource availability at a site (Foster et al. 2003). Thus, it is not just current human activities, but also historical activities that can influence invasion potential.

Some of the major anthropogenic disturbances include urbanization, farming and mining. As part of this work we take one of these disturbances, mining, and assess its impact on invasion. During the surface mining, all vegetation and soils are removed, making it one of the most extreme disturbance events. Post-mining sites cover almost one percent of the world's land (Walker 1999) and represent important land forms in many regions. The invasion potential of post-mining sites depends how they are managed after mining has been completed. The two most common approaches are technical reclamation, comprising of covering the site with topsoil, sowing grass–herb mixtures and planting trees; and spontaneous succession without any human intervention. A third approach, directed succession, when natural processes are actively influenced; e.g., through support of conservation desired plants (by sowing or species-rich hay transferring), or suppressing invasive plants is rarely used (Zipper et al. 2011). By studying such a disturbed environment we will gain further insight into how disturbance is influencing invasions.

Environmental characteristics also play an important role in determining a site's vulnerability to invasion. Prevailing climate, the availability of nutrients and light levels collectively determine establishment. Sites with ample water and nutrients are often the first to be invaded (Thompson et al. 2001). The importance of these habitat characteristics vary not only temporally

with the stage of invasion but also spatially with the geographical scale at which the invasion is being considered. This results in interlaced patterns at multiple spatial and temporal scales (Wagner & Fortin 2005). Spatial scale can be defined at three levels; community, local and regional (Melbourne et al. 2007). This work focuses on local and regional scales, first at a county level, then eco-region and finally the southeastern United States. Once an alien species occurs new regional climate and the broad scale land use (forested, farming, urban) can have a strong influence on the extent of invasion. Within the potential extent of the invasion that the species is found is often driven by small scale characteristics, including microclimate, forest community characteristics and disturbance regime. These influence the persistence of the invasive species. Understanding the role of these variables for invasive alien plants is challenging because environmental variables are often inter-correlated, many invaders have broad ecological niches, and their distributions are often highly dynamic in both space and time (Hulme 2003).

Human induced climate change may intensify the threat of invasive plants (Clements & Ditommaso 2011; Crossman et al. 2011). The innate traits that assist a species in invasion may also help it respond rapidly to changing conditions (Richardson et al. 2000), are the same that may assist in allowing them to respond to shifting niches more rapidly than natives (Dukes & Mooney 1999). Changes that influence plant invasion include rising temperature, altered precipitation, increased atmospheric carbon dioxide, increased nitrogen deposition, and novel disturbances associated with changes in land use. Most of these are expected to increase the risk of invasion.

Land managers need to prioritize areas selected for control to optimize costs and the effectiveness of their control measures. In order to make these difficult decisions, land managers require information on the current distribution and abundance of invasive species and effective techniques to monitor their presence (Byers et al. 2002). They also require predictive models that identify areas that are likely to be invaded. The spread of invasions are influenced by the landscape pattern and scale, thus tools that integrate space, time and scale are essential to understanding the underlying processes. Geographical Information Systems (GIS) are a tool to manage and analyse spatial information, thus a useful instrument in modelling invasive species. The need for integration of both GIS and ecological simulation models is well recognized (Erunova et al. 2006; Gutierrez et al. 2005; Ming & Albrecht 2004; Vakalis 2004).

Species distribution models (SDMs) are now widely used in conservation and ecology (Elith & Leathwick 2009; Franklin 2009). They provide description of the relationship between species

distribution and environmental parameters and are often used to make predictions for unsampled locations (Franklin 2009). SDMs combine concepts from ecology and natural history with more recent developments in statistics and geospatial information systems (Elith & Leathwick 2009) and form the core of predictive geographical modelling in ecology (Guisan & Zimmermann 2000). SDMs build on the species niche concept, defined by Elton (1927) as a species functional role in the biotic community, and refined by Hutchinson (1987) as the hyper-volume as given by the environmental dimensions within which a species can survive and reproduce. SDMs rely on the niche concept that emphasizes species requirements, represented by predominately abiotic factors controlling species distribution (Franklin 2009). Ideally SDMs are developed using comprehensive absence and presence data from the full geographic range for a given species and in locations where the species are near equilibrium with their environment. SDMs can perform well in identifying the natural distributions of species, particularly with ecologically relevant predictors and well-designed survey data that is analysed with an appropriate SDM at the relevant scale (Franklin 2009). In these cases models can provide strong predictive capability and valuable ecological information. However, the strength of models can be limited when species are not at equilibrium with their environment, models are extrapolated in time or space beyond the limit of the data, or when inadequate data are used (Elith & Leathwick 2009). In this work I have used two different SDM approaches, logistic regression and maximum entropy modelling, with similar methods of implementation across chapters.

Logistic regression, a type of generalized linear model, was used to predict the occurrence of invasive plants. Logistic regression has been widely applied in ecology to investigate the relationship between a categorical outcome and a set of explanatory variables or for predicting the probability of occurrence of an event by fitting data to a logistic curve (Hosmer & Lemeshow 2000). It makes use of several predictor variables that may be either numerical or categorical, and here we used it to predict the probability (between 0 and 1) that a species will occur, based on the environmental variables. The goal of this analysis is to find the best fitting and most parsimonious model that is both biologically reasonable and computationally practical for mapping, balancing fit with practicality. In this study logistic regression was applied to those invasive species that occurred in ≥ 50 sampling locations. Data was re-sampled when necessary to give at least a 20% occurrence for each species (Oommen et al. 2010).

Before fitting the logistic regression model an assessment of correlation was performed on independent variables, resulting in the removal of some variables and thus reducing

multicollinearity. Multicollinearity occurs when two or more variables are correlated. Highly correlated (not independent) variables can result in an increased standard error, wider confidence intervals, reduced statistical significance and erratically acting coefficients (Allison, 1999). Any correlation between variables with an $R^2 > |0.5|$, as assessed using Pearson's correlation coefficient, were considered to present a collinearity problem. Collinearity was dealt with by removing the correlated variable that seemed least relevant (either ecologically or one that was hardest to obtain). Although variable exclusion risks bias and a loss of explanatory power (Menard, 2002), it was felt that all correlations between habitat variables with $R^2 >|0.5|$ highlighted genuine repetition in recording of an ecologically relevant parameter. In the later Chapters (Chapter 6 and 7) this was also checked with maximal information coefficients, which assesses nonlinear relationships. A logistic regression was run with all remaining variables. Due to the computation time involved in producing output maps a simple model was desired. Thus backward selection was used to mimic the selection techniques of MaxEnt, with the least significant variable (highest P value) removed from the model and the logistic regression run again. This process was repeated, systematically removing the least significant variable one at a time, until all the variables in the model were significant at the 1% level ($P< 0.01$). Only linear relationships were considered, again this was predominantly to simplify models for mapping. This selection technique may lead to over fitting of the model but for computational simplicity required for mapping, it was the cleanest solution. Test data was used to evaluate the models, thus even if the models were over fitted to the training data, the test data would show in lower assessment statistics. One reason for this approach is that it is similar to the model selection process used in MaxEnt, and there would be some consistency between methods and outputs. This was further checked using AIC, and if there were major variations between AIC based selection and the technique used, I assessed it further. In a few cases there was one variable difference between the two approaches, but it was always the variables that contributed least to the models and did not have a dramatic impact on the model fit. In Chapter 3 we did not use any selection beyond preselecting based on ecological relevance and checking for correlation between variables. In this case we fitted the full model as the size of the study area was limited and a full model could be used in mapping.

Each habitat/environmental group was first analysed separately (i.e. soils, ground, and forest; or Landsat, anthropogenic, climate, environmental, land cover, and water), and the variables showing significance in the separate logistic regressions were used in the final overall model. I

used a lower significance than commonly used (typically $p<.05$) because in many chapters a piecewise (bit by bit) process was undertaken, whereby multiple models were combined, thus the overall significance would be 5% (i.e. five models total, four sub-models, and a final model, the overall p-value of the analysis is limited to 0.05 for each species).

In general the reliability of the model was calculated by using a test and training dataset (30/70). Training data was used to construct the model and test data was used to assess the model's performance. Depending on the chapter, a varying number of replications were run (if there was limited data more replications were run, this was kept consistent with MaxEnt replications) during variable selection. However the final model was based on one run from which the assessment statistics and final maps were produced. The exception to this is Chapter 2 where rather than withholding data, the models were built on the full dataset and the variables selected were then rerun with a selection of the data to assess the stability in the coefficients and model fit.

In most chapters accuracy of prediction was assessed using test AUC and test type II error (FN/(FN+TP; also called omission rate). Type II error was assessed based on a threshold value determined by maximizing specificity and sensitivity (Manel et al. 2002). In some chapters, due to variation in species occurrence across the study area, benchmark type II errors were defined as if data were randomly assigned, and a decrease of more than 25% was considered a useful model. We used the following classes of AUC to assess model performance: 0.50 – 0.75 = fair, 0.75 – 0.92 = good, 0.92 – 0.97 = very good, and 0.97 – 1.00 = excellent (Hosmer & Lemeshow, 2000). In some chapters we also calculated percentage concordance between models, false omission rate (FN/(FN+TN), and used independent datasets. For logistic regression variable importance was calculated using the Wald chi squared statistic, dropping the intercept Wald chi square and standardizing the remainder to 100.

MaxEnt is a machine learning method that compares presence locations to environmental variables at those locations and then across the study area using principles of maximum entropy to generate predictions of suitable habitat in un-sampled regions (Phillips et al., 2006). It is based on the premise that the unknown probability distribution should have maximum entropy, but is constrained by the environmental characteristics of the niche. MaxEnt and other presence only methods are appropriate for modelling species with unstable distributions such as invasive species as true absence data can be difficult to obtain. Absence could be because it has not yet invaded or because the location is unsuitable, and these two options are often indistinguishable

for invasive species (Jarnevich and Reynolds 2011). A drawback of maximum entropy is that it may give very large predicted values for environmental conditions outside the range present in the study area (Phillips et al. 2006). This is due to the exponential modelling process which can lead to values outside of the modelled range being poorly represented, as there is no upper limit to the predicted values. This is not an issue in this study since we are only predicting distribution from within the range that the data has been collected. In MaxEnt the true potential distribution of a species is represented as a probable distribution over the set of cells across the study area. The probability distribution is derived from a set of constraints based on occurrence data. These constraints are stated in terms of functions of environmental variables, called features. Then amongst all probable distributions that meet the constraints the most unconstrained one is used (one of maximum entropy) (Philips and Dudik 2008). MaxEnt controls over fitting and variable selection using a regularization that smooths the modelled distribution, with a penalized maximum likelihood model that balances model fit with model complexity.

Selection of species data followed the same method as for logistic regression, however MaxEnt is based on the occurrence records thus resampling is not necessary. Multicollinearity is not a limitation of MaxEnt, however computational limitation in producing maps makes removal of unnecessary variables essential. These were assessed in with the same methods as for logistic regression. MaxEnt was run with all remaining variables. Due to the computation time involved in producing output maps the simplest model without losing too much power was desired. Thus backward selection was used, with the variables that contributed the least to the model removed until all remaining variables contributed more than 5% to the model, this was based on ten model runs. The key variables in determining the occurrence of each species were identified by their percentage contribution to the final model, and with a jack-knife test on gain and influence on the area under the curve (AUC). This approach assists in reducing a model that over fits.

As with logistic regression, each habitat/environmental group was first analysed separately (i.e. soils, ground, and forest, or Landsat, anthropogenic, climate, environmental, land cover, water) and the variables used in each subgroup then used in the final model. The same assessment statistics were used as with logistic regression, with a modification in how AUC was calculated. The AUC is a test statistic that uses presence and absence records to assess model predictive performance across a range of thresholds. MaxEnt is a presence-only algorithm; therefore we used the Phillips et al. (2006) approach that applied randomly selected pseudo-absences instead of observed absences to AUC. While the MaxEnt model is being trained, it

keeps track of which environmental variables are contributing to fitting the model. Each step of the MaxEnt algorithm increases the gain of the model by modifying the coefficient for a single feature. The program assigns the increase in gain to the environmental variable(s) that the feature depends on (Phillip et al 2006) to give an overall variable contribution to the model.

One discipline within which SDMs have much to offer is in the assessment of the potential distribution of invasive species. Species distribution models have been used with success in the assessment of invasive species (Peterson & Vieglais 2001; Peterson 2003; Andersen et al. 2004) and for predicting the potential distribution of invasive species in regions other than their native ranges. However, model uncertainty presents many challenges. Models necessarily simplify the real world and can ignore important aspects of species ecology (Hampe 2004). This concern with the applicability of SDM to range shifting species or species not at equilibrium has been identified in a number of studies (Elith et al. 2010). One approach is to use an ensemble of predictions based on several different modelling methods (Roura-Pascual et al. 2008; Marmion et al. 2009; Lemke et al. 2011). Others have suggested assessing model behaviour under different data treatments and judging performance based on biological relevance is the best approach (Elith et al. 2010). A number of approaches have been proposed for combining outputs of individual models for ensemble predictions (Araujo & New 2007). In Chapter 3 we assessed two approaches to ensemble predictions. The first is unweighted and averages the logistic and MaxEnt probabilities to derive a combined probability of occupancy between 0 and 1. The threshold probability for predicted presence in the ensemble model is simply the average threshold used in each of the two different models. The second approach is weighted and calculates a new threshold probability as previously used by maximizing the sum of sensitivity and specificity. Where this threshold value departs from the arithmetic average of the two model thresholds, it will weigh the probability scores of one model more than another. In other chapters we adopt a simpler consensus approach for species and climate ensemble predictions, identifying areas of agreement and disagreement in the models based on the binary outputs. The ensemble predictions using different datasets were undertaken in two ways. First, an ensemble approach in which two models outputs were combined. The second was combining the two datasets before modelling.

Until recently, one of the greatest limitations in species distribution modelling for invasive species has been the lack of data (Phillips et al. 2009; Hulme & Weser 2011). Specifically, distribution models are limited by availability of information about the locations where species

are found and the corresponding biological data at high resolution. This has been changing, both in terms of accessibility of location information and an increase in the resolution of data, resulting in a dramatic increase in distribution modelling. Unfortunately, datasets are rarely comprehensive at larger scales and when integrated and supplemented with data at a finer scale there are numerous issues with compatibility (Hulme & Weser 2011). With ease of access to data and now user friendly graphic interface modelling packages, the potential for SDM to be misapplied is growing. A comprehensive description of all independent variables used in this thesis, the resolution and scaling is given in Chapter 9 (Table A9-16). They show the diversity as well as the limitations of current data.

This thesis focuses on applying currently available tools and datasets, assessing their applicability and compare and contrasting their value. I have assessed and modelled invasion of alien plant species to the forest of the southeastern United States. I focus on ten invasive species: Chinese lespedeza (*Lespedeza cuneata*), tall fescue (*Lolium arundinaceum*), Japanese honeysuckle (*Lonicera japonica*), Chinese privet *(Ligustrum sinense*), autumn olive (*Elaeagnus umbellata*), princesstree (*Paulownia tomentosa*), silktree (*Albizia julibrissin)*, chinaberry (*Melia azedarach)*, tree of heaven (*Ailanthus altissima)*, tallowtree (*Triadica sebifera)*. This is done at differing scales, local (Chapter 2 and 3), eco-regional (Chapter 4 and 5) and regionally (Chapters 6 and 7), using field based measurements integrated with remotely sensed and digital datasets and applying both parametric and non-parametric modelling approaches.

1.1 Literature Cited

Allison P.D. (1999) Logistic regression using the SAS systems: theory and application. Cary, NC: SAS Institute.

Andersen, M.C., Adams, H., Hope, B. & Powell, M. (2004) Risk analysis for invasive species: general framework and research needs. Risk Analysis, 24, 893-900.

Araujo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. Trends in Ecology and Evolution, 22, 42–47.

Brown, J.H. & Sax, D.F. (2004) An essay on some topics concerning invasive species. Austral Ecology, 29, 530-536.

Brown, R.L. & Peet, R.K. (2003) Diversity and invasibility of southern Appalachian plant communities. Ecology, 84, 32-39.

Buckley, Y.M., Bolker, B.M. & Rees, M. (2007) Disturbance, invasion and re-invasion: managing the weed-shaped hole in disturbed ecosystems. Ecology Letters, 10, 809-17.

Byers, J.E., Reichard, S., Randall, J.M., Parker, I.M., Smith, C.S., Lonsdale, W.M., Atkinson, I. A. E., Seastedt, T.R., Williamson, M., Chornesky, E. & Hayes, D. (2002) Directing research to reduce the impacts of nonindigenous species. Conservation Biology, 16, 630-640.

Clements, D.R. & Ditommaso, A. (2011) Climate change and weed adaptation: can evolution of invasive plants lead to greater range expansion than forecasted? Weed Research, 51, 227-240.

Crossman, N.D., Bryan, B.A. & Cooke, D.A. (2011) An invasive plant and climate change threat index for weed risk management: Integrating habitat distribution pattern and dispersal process. Ecological Indicators, 11, 183-198.

Davis, M. A., Grime, J.P. & Thompson, K. (2000) Fluctuating resources in plant communities: a general theory of invasibility. Journal of Ecology, 88, 528-534.

Dukes, J.S. & Mooney, H.A. (1999) Does global change increase the success of biological invaders? Trends in Ecology & Evolution, 14, 135-139.

Elith, J. & Leathwick, J.R. (2009) Species distribution models: Ecological explanation and prediction across space and time. Annual Review of Ecology Evolution and Systematics, 40, 677-697.

Elith, J., Kearney, M. & Phillips, S. (2010) The art of modelling range-shifting species. Methods in Ecology and Evolution, 1, 330-342.

Elton, C. (1927) Animal Ecology. Chicago, University of Chicago Press.

Erunova, M., Sadovsky, M. & Gosteva, A. (2006) GIS-aided simulation of spatially distributed environmental processes at “Stolby” state reservation. Ecological Modelling, 195, 296-306.

Fischer, J. & Lindenmayer, D.B. (2007) Landscape modification and habitat fragmentation: a synthesis. Global Ecology and Biogeography, 16, 265-280.

Foster, D., Swanson, F., Aber, J., Burke, I., Brokaw, N., Tilman, D. & Knapp, A. (2003) The importance of land-use legacies to ecology and conservation. BioScience, 53, 77.

Franklin, J. (2009) Mapping species distributions: spatial inference and prediction, 3rd ed. Cambridge University Press, New York.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. Ecological Modelling, 135, 147-186.

Gutierrez, A.P., Pitcairn, M.J., Ellis, C.K., Carruthers, N. & Ghezelbash, R. (2005) Evaluating biological control of yellow starthistle (*Centaurea solstitialis*) in California: A GIS based supply–demand demographic model. Biological Control, 34, 115-131.

Hampe, A. (2004) Bioclimate envelope models: what they detect and what they hide. Global Ecology and Biogeography, 13, 469-471.

Hawkes, C.V., Wren, I.F., Herman, D.J. & Firestone, M.K. (2005) Plant invasion alters nitrogen cycling by modifying the soil nitrifying community. Ecology Letters, 8, 976-985.

Hodkinson, D.J. & Thompson, K. (1997) Plant dispersal: the role of man. Journal of Applied Ecology, 34, 1484-1496.

Holmes, T.P., Aukema, J.E., Holle, B. Von, Liebhold, A. & Sills, E. (2009) Economic impacts of invasive species in forests. Annals of The New York Academy of Sciences, 1162, 18-38.

Hosmer, D.W. & Lemeshow, S. (2000) Applied Logistic Regression. Wiley-Interscience, New York.

Hulme, P.E. & Weser, C. (2011) Mixed messages from multiple information sources on invasive species: a case of too much of a good thing? Diversity and Distributions, 7, 1152–1160.

Hulme, P.E. (2003) Biological invasions: winning the science battles but losing the conservation war? Oryx, 37, 178-193.

Hulme, P.E., Pyšek, P., Nentwig, W. & Vilà, M. (2009) Will threat of biological invasions unite the European Union. Science, 324, 40–41.

Hutchinson, M.F. (1987) Methods for generation of weather sequences. In Bunting, A.H. (Ed.) Agricultural environments: characterisation, classification and mapping. Wallingford, CAB, International, 149-157.

Jarnevich, C. & Reynolds, L. (2011) Challenges of predicting the potential distribution of a slow-spreading invader: a habitat suitability map for an invasive riparian tree. Biological Invasions, 13, 153-163.

Lemke, D., Hulme, P.E., Brown, J.A. & Tadesse, W. (2011) Distribution modelling of Japanese honeysuckle (*Lonicera japonica*) invasion in the Cumberland Plateau and Mountain Region, USA. Forest Ecology and Management, 262, 139-149.

Manel, S., Williams, H.C. & Ormerod. S.J. (2002) Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology, 38, 921-931.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009) Evaluation of consensus methods in predictive species distribution modelling. Diversity and Distributions, 15, 59-69.

Martin, M.R., Tipping, P.W. & Sickman, J.O. (2009) Invasion by an exotic tree alters above and belowground ecosystem components. Biological Invasions, 11, 1883-1894.

Melbourne, B. a, Cornell, H.V., Davies, K.F., Dugaw, C.J., Elmendorf, S., Freestone, A.L., Hall, R.J., Harrison, S., Hastings, A., Holland, M., Holyoak, M., Lambrinos, J., Moore, K. & Yokomizo, H. (2007) Invasion in a heterogeneous world: resistance, coexistence or hostile takeover? Ecology letters, 10, 77-94.

Menard, S.W. (2002) Applied logistic regression analysis. Sage Publications, Inc

Miller, J.H., Chambliss, E.B. & Loewenstein, N.J. (2010) A field guide for the identification of invasive plants in southern forests. Gen. Tech. Rep. SRS–119. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 126 p.

Ming, P. & Albrecht, J. (2004) Integrated framework for the simulation of biological invasions in a heterogeneous landscape. Transactions in GIS, 8, 309-334.

Moser, W.K., Barnard, E.L., Billings, R.F., Crocker, S.J., Dix, M.E., Gray, A.N., Ice, G.G., Kim, M., Reid, R., Rodman, S.U. & McWilliams, W.H. (2009) Impacts of nonnative invasive species on US forests and recommendations for policy and management. Journal of Forestry, 107, 320-327.

Oommen, T., Baise, L.G. & Vogel, R.M. (2010) Sampling Bias and Class Imbalance in Maximum-likelihood Logistic Regression. Mathematical Geosciences, 43, 99-120.

Peterson, A.T. & Vieglais, D.A. (2001) Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. BioScience, 51, 363.

Peterson, A.T. (2003) Predicting the geography of species invasions via ecological niche modeling. Quarterly Review of Biology, 78, 419-433.

Pfeiffer, J.M. & Voeks, R.A. (2008) Biological invasions and biocultural diversity: linking ecological and cultural systems. Environmental Conservation, 35, 281-293.

Phillips, S. J. and Dudík, M. (2008), Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography, 31: 161–175.

Phillips, S., Anderson, R. & Schapire, R. (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190, 231-259.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications, 19, 181-197.

Pimentel, D., Zuniga, R. & Morrison, D. (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecological Economics, 52, 273-288.

Pysek, P. & Richardson, D.M. (2007) Traits Associated with Invasiveness in Alien Plants: Where Do we Stand? In Biological Invasions, Ecological Studies (Vol. 193) (Nentwig, W., ed.), pp. 97–125, Springer-Verla

Pyšek, P. & Richardson, D.M. (2010) Invasive Species, Environmental Change and Management, and Health. Annual Review of Environment and Resources, 35, 25-55.

Pysek, P. (2002) Patterns of invasion in temperate nature reserves. Biological Conservation, 104, 13-24.

Reinhart, K.O., Greene, E. & Callaway, R.M. (2005) Effects of *Acer platanoides* invasion on understory plant communities and tree regeneration in the northern Rocky Mountains. Ecography, 28, 573-582.

Ricciardi, A. (2007) Are modern biological invasions an unprecedented form of global change? Conservation Biology, 21, 329-36.

Richardson, D.M., Allsopp, N., D'Antonio, C.M., Milton, S.J. & Rejmánek, M. (2000) Plant invasions--the role of mutualisms. Biological reviews of the Cambridge Philosophical Society, 75, 65-93.

Roura-Pascual, N., Brotons, L., Peterson, A.T. & Thuiller, W. (2008) Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. Biological Invasions, 11, 1017-1031.

Standish, R., Robertson, A. & Williams, P. (2001) The impact of an invasive weed *Tradescantia fluminensis* on native forest regeneration. Journal of Applied Ecology, 38, 1253-1263.

Strauss, S.Y., Lau, J. A. & Carroll, S.P. (2006) Evolutionary responses of natives to introduced species: what do introductions tell us about natural communities? Ecology Letters, 9, 357-74.

Theoharides, K.A. & Dukes, J.S. (2007) Plant invasion across space and time: factors affecting nonindigenous species success during four stages of invasion. New Phytologist, 176, 256-273.

Thompson, K., Hodgson, J.G., Grime, J.P. & Burke, M.J.W. (2001) Plant traits and temporal scale: evidence from a 5-year invasion experiment using native species. Journal of Ecology, 89, 1054-1060.

Vakalis, D. (2004) A GIS based operational system for wildland fire crisis management I. Mathematical modelling and simulation. Applied Mathematical Modelling, 28, 389-410.

Vitousek, P.M., D'Antonio, C.M., Loope, L.L., Rejmanek, M. & Westbrooks, R. (1997) Introduced species: A significant component of human-caused global change. New Zealand Journal of Ecology, 21, 1-16.

Wagner, H.H. & Fortin, M.J. (2005) Spatial analysis of landscapes: concepts and statistics. Ecology, 86, 1975-1987.

Walker, L.R. (1999) Ecosystems of disturbed ground. Ecosystems of the world 16. Elsevier, Amsterdam.

Webster, C.R., Jenkins, M.A. & Jose, S. (2006) Woody invaders and the challenges they pose to forest ecosystems in the eastern United States. Journal of Forestry, 104, 366.

Wilcove, D.S., Rothstein, D., Dubow, J., Phillips, A. & Losos, E. (1998) Quantifying threats to imperilled species in the United States. BioScience, 48, 607.

Zipper, C.E., Burger, J.A., McGrath, J., Rodrigue, J.A. & Holtzman, G.I. (2011) Forest restoration potential of coal-mined lands in the eastern United States. Journal of Environmental Quality, 40, 1567-1577.

CHAPTER 2.

Invasion of a Mined Landscape: what habitat characteristics influence the occurrence of invasive plants?

Collaborators: *C. J. Schweitzer, I. A. Tazisong, and Y. Wang*

2.1 Abstract

Throughout the world, the invasion of alien plants is an increasing threat to native biodiversity. Invasion is especially prevalent in areas affected by land transformation and anthropogenic disturbance. Surface mines are a major disturbance, and thus may promote the establishment and expansion of invasive plant communities. Using the Shale Hills region of Alabama as a case study, we examined the environmental and habitat factors that may contribute to favourable conditions for heightened plant invasion, and developed models for predicting the probability of occurrence of invasive plant species. We conducted vegetation surveys, soil sampling, and environmental evaluation on the mined landscape. Canonical correspondence analysis suggested that the invasive community was predominantly associated with forest structure and composition. Chinese privet (*Ligustrum sinense*) and Japanese honeysuckle (*Lonicera japonica*) were more likely to be present in areas with higher vegetation diversity, total canopy cover and hardwood density; autumn olive (*Elaeagnus umbellata*) and princesstree (*Paulownia tomentosa*) were more likely to be present in areas with higher hardwood basal area. These parameters are features often associated with more established forests. Chinese lespedeza (*Lespedeza cuneata*) and shrubby lespedeza (*Lespedeza bicolor*) were less likely to be found at sites with the above characteristics. Logistic regressions with the three most common species, Chinese privet, Japanese honeysuckle and Chinese lespedeza, all had reasonable concordance (>75) and over 25 percent decrease in false omission rates and type II errors, suggesting useful models for predicting occurrence. Chinese lespedeza is more likely to be found in open or pine areas with higher magnesium levels in the soil and little or no midstory and downed woody debris. Japanese honeysuckle was more likely found in areas with high canopy cover with little

midstory and in areas of high soil magnesium and higher species diversity. Chinese privet was more likely found in areas with high canopy cover. The influence of planting alien, invasive species in this area is likely the major driver of the high diversity of invasive plants, with three of the six dominant species being planted. Adjusting the reclamation plantings to native species would aid in resolving this. Overall, it appears that the initial reclamation efforts, apart from the planting of invasive species, are not the major driver impacting the alien, invasive species composition of the reclaimed, now forested mine sites.

2.2 Introduction

Land transformation and anthropogenic disturbance often facilitate the establishment and development of invasive plant community. Surface mining is one of the major forms of disturbance and has changed over 2.4 million hectares of terrestrial habitat in the United States since 1930 (Zeleznik & Skousen 1996). The changes include alteration in ecosystems and geophysical characteristics (Holl 2002; McSweeney & Jansen 1984; Negley & Eshleman 2006; Shukla et al. 2005). Impacts include interruption and change of energy flow, food webs, biodiversity, successional patterns, and biogeochemical cycling (Ripley et al. 1996). Surface mining is distinct from most other land disturbances in that the disturbance is comprehensive, with native vegetation, soils, soil microbes, and seed banks being removed.

Since the introduction of the Surface Mining Control and Reclamation Act (SMCRA) in 1977, much of the land transformed by surface coal mining in the United States has been subjected to some reclamation, with efforts aimed at improving the quality of the land by restoring some of the pre-disturbance vegetation and functions (Bradshaw 1984). Reclamation starts before the mining operation, with each mine needing an approved reclamation plan before commencement of operation. The first stage in the mining operation is normally the removal of the top stratum. Often this is a very shallow layer of unconsolidated rubble and soil that is retained for use in the reclamation. In most cases, topsoils are not present and a heterogeneous mixture of suitable overburden materials from this top stratum is used as the final growth medium in the reclamation. On completion of the coal removal the surface mine site is re-contoured and stabilized, covered with ‘topsoil’ and then vegetated. Surface mine reclamation efforts rarely result in ecosystems that mimic pre-mined characteristics; the focus is generally on short-term measurable matrices including land stability and hydrological function. However, in

recent years, in the eastern United States there has been growing interest in the restoration of forest community, structure and function (Zipper et al. 2011).

Throughout the world, alien plants are becoming an increasing threat to native biodiversity and ecosystem functions (Ricciardi 2007; Vitousek et al. 1997). Historically and still to some extent today, alien species are used in reclamation, to stabilize land and quickly develop a vegetation community. In disturbed systems such as mined areas, non-native invasive plants can be a significant management concern reducing ecosystem services. Invasive plants can change ecosystem services and influence the long-term ecological and economic productivity of land (Webster et al. 2006). Invasiveness (traits that enable a species to invade a new habitat) and invasibility (the susceptibility of a community or habitat to the establishment and spread of new species) are key components for the occurrence and spread of alien plants (Alpert et al. 2000). The characteristics of plants that assist in some of the short-term goals of restoration, including land stabilization and nitrogen fixing are often the same traits that are associated with invasive plants. Some of the traits reclamationists favour in their choice of plants, including fast establishment, the ability to grow under harsh conditions, and adaptation to nutrient-poor soils also relate to invasive tendencies. Habitat attributes that are associated with invasibility are disturbance, early successional environments, low diversity of native species (Lodge 1993), and high environmental stress (Alpert et al. 2000; D'Antonio et al. 1999; Skousen et al. 1994). Mined sites often display these attributes and thus may have a high probability of being invaded by unwanted species.

In the southeastern region of the United States, the counties with the highest diversity of invasive plants occur in the Southern Piedmont, Interior Low Plateau, and Southern Ridge and Valley of the Appalachian-Cumberland highlands (Miller et al. 2012). The same report showed that highest density of survey points with invasive plants (purple in Figure 2-1) was in the top half of Alabama . These areas have had a long history of human habitation and highly disturbed mining regions.

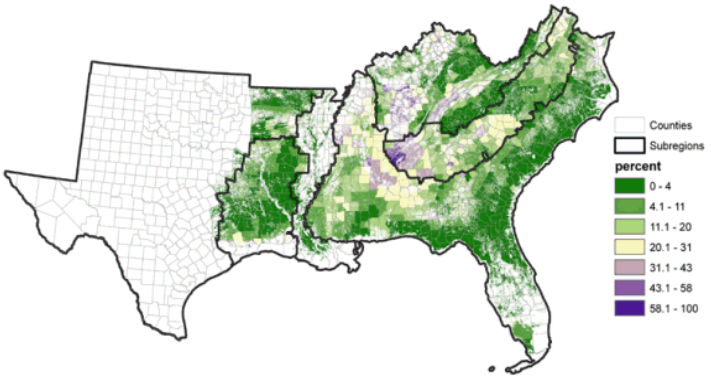

Figure 2-1: Percent of survey plots within a county occupied by one to four invasive plants, 2010 (Source: Forest Inventory and Analysis, Southern Research Station, U.S. Forest Service) from (Miller, Lemke, & Coulston 2011).

We investigated the occurrence of invasive plants in the Shale Hills Region (SHR) in mid-Alabama, quantified habitat and environmental conditions, examined the associations of invasive community and habitat and ecological characteristics, and developed predictive models for the occurrence of invasive species.

2.3 Methods

2.3.1 Study Area

Our study was conducted in the Shale Hills Region (SHR) of the southern Cumberland Plateau of the south-eastern United States (Figure 2-2). The southern Cumberland Plateau has a temperate climate characterized by long, moderately hot summers and short, mild winters (Smalley 1979). The average minimum winter temperature is 1 °C, and the average summer maximum temperature is 32 °C (Smalley 1979). Annual precipitation averages approximately 1400 mm and is fairly well distributed throughout the year (Smalley 1979). Precipitation is greatest from January through April and least from August through November (Smalley 1979). Thunderstorms with high intensity rainfall are common in the summer (Smalley 1979). The forests of the Cumberland Plateau are among the most diverse of the world's temperate-zone forests (Ricketts et al. 1999). Like much of the forests in the eastern United States, the native deciduous hardwood and mixed pine hardwood ecosystems of the Cumberland Plateau have undergone a long history of land-use change (McGrath et al. 2004; Wear & Greis 2002). This area has undergone extensive land use changes, including surface mining, that have altered the landscape and ecosystem functions. The SHR comprises the southern extremity of the Cumberland Plateau. Topography is rugged and fairly complex. Because ridge tops are much lower than those in northern sections of the Plateau, the characteristics of the sub-region is one of extensive hills, not mountains or a plateau. Strongly sloping land predominates, and the area is

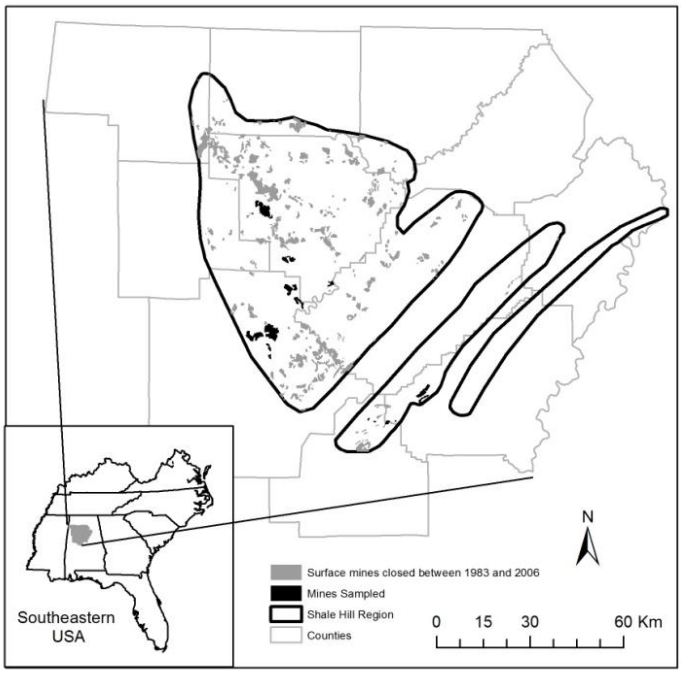

Figure 2-2: Study area location map, Shale Hills region, Alabama.

mostly forested. In this area, dissection has largely removed the parent soil's sandstone cap and exposed the underlying shale. Coal mining, both shaft and strip, is a major industry (Smalley 1979). Our target area included surface mines permitted after 1983, on both public and private lands, that were closed before 2006, thus had time to be reclaimed and for vegetation to re-establish. The final phase of restoration is planting of the permanent vegetation. Mines considered in this study were planted at a rate of 500 to 700 pines per acre, with 450 pines per acre considered successful (Skelly and Loy 1979).

2.3.2 Species of Interest

The study area has many of the 56 alien plants that are highly invasive to the forests of the south-eastern United States (Miller et al. 2010). In this study we focused on the six most prevalent species: shrubby lespedeza (*Lespedeza bicolor*) (found at 20 sites (n=20)), Chinese lespedeza (*Lespedeza cuneata*) (n=300), Japanese honeysuckle (*Lonicera japonica*) (n=217), Chinese privet (*Ligustrum sinense*) (n=68), autumn olive (*Elaeagnus umbellata*) (n=29), and princesstree (*Paulownia tomentosa*) (n=22). Following are brief descriptions of each of these six species.

2.3.2.1 Shrubby lespedeza (*Lespedeza bicolor*)

Shrubby lespedeza was introduced from Japan in the 1800s as an ornamental. It has been planted for wildlife habitat (Davison 1945; Haugen & Fitch 1955), and is also used in strip mine reclamation and along field borders (Graham 1941). It can reach three metres in height (Evans et al. 2006) and grows well in open areas, particularly on well-drained and acidic soils (Sun et al. 2008). Shrubby lespedeza is a serious invader in the southeastern region of the United States and is found in 27 states (USDA 2011) throughout the country. It has been planted as part of reclamation in this area since the 1970s (pers com Dr. Randall Johnson, Director, Alabama Surface Mining Commission [ASMC]).

2.3.2.2 Chinese lespedeza (*Lespedeza cuneata*)

Introduced from Japan in 1899, Chinese lespedeza, also called Sericea lespedeza, is a long, slender perennial legume that can grow up to two metres tall. The species has spread quickly due to its use in pasture and erosion control (Miller et al. 2010), along roadways, on reclaimed mines, and along field borders (Graham 1941). It is flood tolerant and can survive in a wide variety of

habitats, including forests, road sides, and open fields (Miller et al. 2010). Chinese lespedeza is found in 31 states in central and eastern United States (USDA 2011). It forms thick clusters that can spread over large areas and ultimately prevent forest regeneration, with seed pods that can stay viable for years (Miller et al. 2010). It has been planted as a part of reclamation in this area since the 1970s (pers com Dr. Randall Johnson, Director, ASMC).

2.3.2.3 Japanese honeysuckle (*Lonicera japonica*)

Japanese honeysuckle is native to Asia (Ohwi 1965) and was introduced to the United States in 1806 (Leatherman 1955), with the first noted escape from cultivation occurring in 1882 (U.S. National Herbarium). It was later widely planted for deer forage (Dickson et al. 1978; Patterson 1976) and is now considered naturalized in upland and lowland forests as well as in forest-edge habitats (Patterson 1976; Yates et al. 2004). It has been documented in at least 42 states within the United States, is listed as an invasive or noxious weed in several eastern states (USDA 2011), and is the most prevalent invasive plant in southeastern forest (Miller et al. 2012). The species occurs in both open and shaded areas, with annual precipitation in invaded areas averaging 1000–1200 mm and minimum temperatures as low as -15 to -8 ^{o}C (Sasek & Strain 1990). Based on the current distribution in United States, its ecology, physiology, and phenotypic plasticity, the species is expected to continue to spread in eastern North America (Schierenbeck 2004). Although it is considered a widespread, naturalized weed, as recently as 1994, it was recommended by wildlife managers for use as deer forage and cover (Dyess et al. 1994).

2.3.2.4 Chinese privet (*Ligustrum sinense*)

Chinese privet was introduced in the 1800s as a decorative shrub (Miller et al. 2010) and is now the most common invasive privet in the southeastern United States, occurring in 20 states, ranging from Texas to Massachusetts (USDA 2011). An evergreen thicket-forming shrub native to China and Europe, the species can grow up to 10 m tall (Miller et al. 2010).

Privet is the second most abundant invasive plant in the South and is most prevalent in the understory of bottomland hardwood forests (Merriam & Feil 2002; Miller et al. 2012). The invasion by this species severely alters natural habitat and critical wetland processes, forming dense stands that exclude most native plants and preventing natural forest regeneration. The abundance of specialist birds and diversity of native plants and bees can be reduced by privet thickets (Hanula et al. 2009; Wilcox & Beck 2007). Privet can survive in a variety of habitats,

including wet or dry areas, but it dominates in mesic forests. Privet produces abundant seeds that are viable for about a year (Shelton & Cain 2002) and are predominately spread by birds (Greenburg & Walter 2010). The species also increases in density by stem and root sprouts. Although controlling privet infestations costs the United States billions of dollars each year (Simberloff et al. 1997), it is still being produced, sold, and planted as an ornamental.

2.3.2.5 Autumn olive (*Elaeagnus umbellata*)

Brought to the United States in 1830 from Japan and China, autumn olive was primarily used for mine reclamation, field rows for erosion control, and wildlife habitats (Miller et al. 2010). Since then it has escaped from cultivation and is now found in 37 states including Hawaii (USDA 2011). Autumn olive can grow in acidic, loamy soils and produces numerous seeds (Travis & Wilterding 2005), it is a nitrogen fixer, thus can do well on poor soils (Sharp 1977). Autumn olive can aggressively colonize an area, once established, it can develop intense shade which suppresses native species, particularly those that flourish on nitrogen-poor soils (Sather & Eckardt 1987). Management is required to contain the spread of this species (Travis & Wilterding 2005), but control by cutting, burning, or the combination is counter-effective and stimulates sprouting and growth (Donovan et al. 2007). It has been planted as part of reclamation in this area since the 1970s (pers com Dr. Randall Johnson, Director, ASMC).

2.3.2.6 Princesstree (*Paulownia tomentosa*)

Native to East Asia, princesstree was introduced into the eastern United States in early 1800s (Miller et al. 2010), and is now found in 25 states in the east and south (USDA 2011). It is still widely sold and planted as an "instant" shade tree. Until recently, most research on princesstree in the United States focused on increasing growth in plantations due to the exceptional timber value in exports to Japan (Johnson et al. 2003; Miller et al. 2010). In the northeast United States, princesstree plantations can produce valuable high quality wood, but in the southeastern region, due to the more favourable growing season, tree growth is too fast, producing low-density wood that is of much lower quality and value. The presence of princesstree is associated with natural disturbance (Williams 1993) and is therefore likely to be promoted by anthropogenic disturbance. Williams (1993) classified the species as a non-aggressive species, though others (Langdon & Johnson 1994) suggested that in areas of high disturbance it shows invasive traits. Although sun-adapted and capable of extremely rapid growth in high light environments,

princesstree is tolerant of a wide range of light levels (Longbrake & McCarthy 2001). Forest management practices can affect the establishment and development of this species with growth and survival on clearcuts being greater than in forest edges or in undisturbed forest (Longbrake 2001).

2.3.3 Sampling Point Selection

Sampling points were selected using the stratified spatial balanced sampling design, Generalized Random Tessellation Stratified (GRTS) (Stevens & Olsen 2004). GRTS design allows flexibility in sampling; the selected sample points are spatially balanced, so that if a point is inaccessible (land access permit and difficult physical conditions), the next point in the sample-list can be selected while maintaining spatial balance. Sampling was also allowed to be extended beyond the initial plan if time permitted while maintaining spatial balance. Two hundred sites were located across the study area with the goal of surveying at least one hundred sites. Site selection was stratified by years since reclamation: >20 years, 10 – 20 years, and <10 years. At each sample site, an adaptive cluster sampling design was used to assess the magnitude of invasive plants and habitat and environmental conditions which might encourage introduction and spread of invasive plant species. Adaptive sampling was employed when individuals of invasive species were found on the main survey plot; four additional sampling plots were used gain more information about the species preferences. As invasive plants are often a rare or clustered event, this approach allows for greater efficiency of research resources by ensuring effort is targeted to where the plants are located (Brown 2003; Kriticos et al. 2003).

2.3.4 Field Sampling

Field sampling occurred from June through October 2010. We sampled 112, 405 m^2 (1/10-acre) circular plots. GPS coordinates, date, time, forest type (pine, mixed or hardwood), regeneration type (natural or planted), distance to established forest, and forest age were recorded on each plot. All trees with ≥ 25 mm diameter at breast height (DBH, ca 1.4 m above ground level) were recorded for species and categorical DBH (25 – 75 mm, 75 – 150 mm, 150 – 225 mm, 225 – 375 mm, or > 375 mm) to assess habitat structural diversity. These categorical groupings were later reduced to three, small (DBH 25 – 75 mm), medium (75 – 225 mm) and large (> 225 mm). An increment borer was used to obtain a tree core from the largest accessible tree in each plot. Two circular subplots of 1.8 m radius were established 3.7 m north and south of

the main plot centre for assessing percentages of overstory, midstory, and understory cover (0 – 1 m) (USDA Forest Service 1998) and the dominant species in each stratum. Ground variables were recorded at each subplot as percent cover of rock, bare soil, litter (tree and grass litter were estimated separately), non-vascular plants and fungi, and downed woody debris. A hand-held spherical densitometer was used to determine the cover of the forest canopy within each of these subplot, two readings were taken at each subplot to give four readings per plot. Leaf litter and humus depth were measured to the nearest mm at north and south edge of each subplot (four readings per plot). After removing the leaf litter from the soil surface, soil samples were taken with a hand-held-probe from 0 – 10 cm depth at the centre of each subplot (two soil samples per plot). The soil samples were air-dried, ground, and sieved using a two mm stainless steel sieve into plastic bags and stored until soil analysis was undertaken. If any invasive plant species was detected, an additional four neighbouring sampling plots, referred to as adaptive plots, were measured, using the same sampling techniques as for the main plot, with the plot centre 33.5 m in each cardinal direction from the main plot centre. In a few cases, it was not possible to reach the additional plot due to water or topography (cliffs); in such cases no data were recorded for that additional plot.

2.3.5 Soil Analysis

Soil pH was measured in water at a soil to solution ratio of 1:2. The pH reported was temperature compensated at 25 ^{o}C. Total C, N, and S in the soil were determined using the dry combustion method with a vario Max CNS analyser (Elementar, Hanau, Germany). Cation exchange capacity (CEC) was measured using the ammonium acetate (pH 7) method. Available micronutrients (Fe, Zn, Cu and Mn) were extracted using DTPA method (Lindsay & Norvell 1978), while macronutrients (K, Ca, Mg, P, Na) were extracted using Mehlich 3 solution (Mehlich 1984) and analysed using inductive couple plasma spectroscopy (ICP, Perkin Elmer, Massachusetts, USA). Inorganic ammonium and nitrate content in the soil were extracted with 2 *M* KCl and analysed using ammonium-nitrate analyser (Timberline Instrument, Model no. TL-2800). Ammonium acetate extractable bases (K, Na, Ca, and Mg) were used to determine percent base saturation of the soil. Once analysis was complete, results were combined for each main plot and used to represent the main plot and surrounding adaptive plots

2.3.6 Analysis of Data

Habitat data were analysed in three groups: soil characteristics, ground variables (from soil to understory), and forest structure (above understory) (Table 2-1). Soils nutrient variables were standardized to a concentration of parts per million (ppm). Ground variables included categorical ground cover recorded as percent, percent understory cover, litter depth and humus depth. Forest structure was estimated using tree measurements and included diversity indices (Shannon & Weaver 1949; Simpson 1949), basal area (of trees with a DBH > 150 mm), and tree density, percent upper and midstory cover, and overall canopy cover. These calculations were conducted for all forest types combined, and then for the pines and hardwoods, separately. Correlations among variables within each habitat group were assessed in SAS (SAS Institute 2009) to exclude the variables with high correlation ($r^2 > 0.50$) from further analysis (Tables A9-1 to A9-4). The selection among highly correlated variables was based on the relative easiness for field application. All uncorrelated variables were tested for any underlying spatial autocorrelation in their structure that may relate more to spatial patterns than the ecological relevance of variables. We used the Mantel test to measure spatial dependence among the samples (Mantel & Valand 1970). If any variables had an r^2 greater than 0.1 they were explored further to assess the impact of spatial autocorrelation on the analysis.

The relationship between habitat variables and the invasive community was initially assessed using Canonical Correspondence Analysis (CCA) in PC-Ord. Invasive plant species that were observed at less than five percent of the sites were excluded for CCA (Heikkinen 1996; Hill 1991). We first assessed the relationship between the invasive community and each of the three groups of habitat variables separately. An overall CCA was then conducted, using the variables that had the strongest associations ($r^2 > 0.30$) based on the three habitat group CCA.

Logistic regression in SAS (SAS Institute 2009) was used to build occurrence predictive models. Logistic regression is a generalized linear model that is used to investigate the relationship between a categorical outcome and a set of explanatory variables or for predicting the probability of occurrence of an event, presence of invasive species in this study, by fitting data to a logistic curve (Hosmer & Lemeshow 2000). As with CCA, each habitat group was first analysed separately (soils, ground, and forest), and the variables showing significance in the separate logistic regression were used in the final overall model. Logistic regression was applied to those invasive species that in occurred in ≥ 50 sampling plots to assure balance in the number of absences and presences (suggested ratio 2:8) in the data (Oommen et al. 2010). A piecewise,

stepwise procedure was used to build the most of the parsimonious model with a p-value of 0.01 for entering or dropping out of model. A p-value of 0.01 was used for each model. With five models total, three sub-models, a combined model and a final model, the overall p-value of the analysis is limited to 0.05 for each species. For descriptive purposes variable importance and direction of relationship were tabulated. Variable importance was calculated using the Wald chi squared statistic, dropping the intercept Wald chi square and standardizing the remainder to 100. Accuracy of prediction was assessed using percentage concordance, false omission rate [(FN/(FN+TN)), FN false negative, TN true negative], and Type II error [(FN/(FN+TP)), TP true positive]. False omission rate and Type II error were assessed based on a threshold value determined by maximizing specificity plus sensitivity (Manel et al. 2001). Due to variation in species occurrence across the study area a benchmark omission rate and type II error were defined as if data were randomly assigned, and a decrease of more than 25 percent was considered a useful model (Hair et al. 2006).

The stability of final models for each species was assessed by re-sampling the data, without replacement. One hundred observations were randomly selected by maintaining the observed occurrence/non-occurrence ratio of that species. A total of 1,000 re-sampling were conducted. If the mean p value of a variable from the re-samples was greater than 0.15, the variable was dropped (Nilsson & Belyaev 1998). It is expected that these models have weaker relationships as the number of data points has been substantially reduced, thus a higher p-value has been used. Standard deviation and 99 percent confidence limits were calculated for the each variable in the final model based on re-sampling runs.

2.4 Results

We sampled a total 374 plots, 112 main plots and 262 adaptive plots. Average age of largest tree was 13 ± standard deviation of 7 years. The ground cover was variable, though predominantly litter in composition, averaging 63 ± 32 percent litter coverage. The predominant herbaceous species was Chinese lespedeza. Understory cover was high at 59 ± 26 percent with midstory averaging 23 ± 20 percent. The sites varied in forest composition from no tree cover to even-aged pine stands to mixed-species of varying ages. Basal area across all sampling plots averaged 43 ± 53 m^2ha^{-1}. Pine was the major component (95% of the total basal area); this is the species of choice when reforesting reclaimed mines in this area. The soils were mostly acidic, with pH ranging from 3.89 to 7.12. Macro and micronutrients content ranged from 1.8

(phosphorus) to 2,468 (calcium) and from 0.3 (copper) to 447 (iron) mg kg^{-1} soil, respectively. The cation exchange capacity ranged from 2.5 to 21.5 cmole kg^{-1} soil, while the percent of base saturation ranged from 2 to 137 percent (Table 2-1). Spatial autocorrelation as measured by Mantels test was low; all variables have r^2 of less than 0.01.

Table 2-1: Habitat variables measured at each sampling plot (X identifies variables with low Pearson's correlations that are used for further analysis)

	Unit	$r^2 < 0.50$	mean	SD	min	max
pH		X	5.55	0.70	3.89	7.12
Phosphorus	ppm	X	10.2	6.5	1.8	34.9
Potassium	ppm	X	163	86	14	440
Sodium	ppm	X	36	16	6	104
Magnesium	ppm	X	249	165	16	746
Calcium	ppm		814	630	42	2468
Iron	ppm	X	192	93	16	447
Zinc	ppm	X	5.4	4.4	0.5	23.7
Copper	ppm	X	2.8	2.1	0.3	10.7
Manganese	ppm	X	99	65	5	340
Calcium Magnesium ratio		X	4.6	7.3	0.5	54.8
Ammonium	ppm	X	11.4	5.8	2.9	43.2
Nitrate	ppm	X	6.9	7.3	0	36.4
% Carbon	%		2.0	1.4	0.1	6.2
% Nitrogen	%	X	0.13	0.08	0.01	0.37
% Sulphur	%	X	0.06	0.09	0.00	0.48
Carbon Nitrogen ratio		X	14.7	4.7	5.8	25.7
cation exchange capacity		X	11.6	3.6	2.5	21.5
% base saturation	%		45	30	2	137
% understory	%	X	59	26	0	100
% Rock	%	X	4	9	0	70
% Bare Soil	%	X	9	15	0	80
% non-vascular plants	%	X	3	6	0	40
DWD	%	X	8	12	0	80
% Shale	%	X	6	14	0	88
% Leaf Litter	%		51	39	0	100
% Grass Litter	%	X	13	21	0	95
%Total Litter	%	X	63	32	0	115
Litter Depth	cm	X	1.8	1.3	0	8.0
Humus depth	cm	X	0.8	1.0	0	5.6
Richness			5	4	0	23
Shannon			0.76	0.67	0	2.61
Simpson's Evenness		X	0.39	0.30	0	1
Hardwood Richness			3	4	0	21

Table 2-1 continued: Habitat variables measured at each sampling plot

	Unit	r^2<0.50	mean	SD	min	max
Oak Richness			0.4	1.2	0	7
Densiometer	%	X	49	33	0	96
% Overstory	%		26	29	0	95
% Midstory	%	X	23	26	0	100
Number of stems per plot		X	52	54	0	388
# of small stems 25 to 75 mm			28	36	0	242
# of medium stems 75 to 225 mm			20	24	0	167
# of large stems greater than 225 mm			3	5	0	34
Basal area of trees greater than 150 mm	m^2/ha	X	43	53	0	303
# of pine stems			36	45	0	388
# of small pine stems 25 to 75 mm		X	17	29	0	235
# of medium pine stems 75 to 225 mm		X	17	23	0	167
# of large pine stems > 225 mm			3	5	0	34
Basal area of pine trees > 150 mm	m^2/ha		39	52	0	301
# of hardwood stems		X	15	32	0	230
# of hardwood stems 25 to 75 mm			11	23	0	203
# of hardwood stems 75 to 225 mm			4	9	0	65
# of hardwood stems > 225 mm			0.4	1.3	0	10
Basal area of hardwood trees > 150 mm	m^2/ha	X	5	15	0	104
Number of heavy seeding hardwood stems			3	11	0	131
Basal area of heavy seeding hardwood trees > 150 mm			2	9	0	86
Forest Age	years	X	13	7	0	50

The CCA of soil variables with the invasive plant community illustrated that autumn olive and princesstree are associated with sites with higher nitrogen, and lower calcium to magnesium ratio; Chinese privet and Japanese honeysuckle were associated with high manganese; whereas Chinese and shrubby lespedeza were associated with lower nitrogen, and higher calcium to magnesium ratio (Figure 2-3a). The first two CCA axes with soil features were weak, explaining only 12 percent variation within the invasive community (Figure 2-3a). The CCA of ground variables with the invasive plant community showed shrubby lespedeza and princesstree preferred sites with more bare soil, Chinese lespedeza was associated with high grass litter cover; Japanese honeysuckle and Chinese privet were strongly associated with litter depth and litter cover; and autumn olive was most strongly associated with downed woody debris (Figure 2-3b). The first two CCA axes of ground variables were also weak, explaining only 9 percent variation within the invasive community (Figure 2-3b). The CCA of forest variables with the invasive plant community showed stronger associations including hardwood basal area with princesstree

and autumn olive, along with high canopy cover, high diversity and high hardwood density with Japanese honeysuckle and Chinese privet; Chinese and shrubby lespedezas were negatively associated with the forest structure variables (Figure 2-3c). The first two CCA axes of forest structure variables explained 18 percent variation within the invasive community (Figure 2-3c). The first two axes of CCA with selected variables combined from three habitat variable sets explained 19 percent of the invasive community variation (Figure 2-3d). Overall, forest structure variables had the only strong correlations with the invasive plant community; and followed the same pattern as with the forest CCA.

We applied logistic regression to three invasive species that occurred at 50 or more sampling sites: Chinese lespedeza, Japanese honeysuckle, and Chinese privet, using habitat variables selected with limited correlation (Tables 2-2 to 2-4). The regressions that used soils data included seven soils variables (Table 2-3), with no variable dominating all the models. Regression models for the ground component used six of the eight variables, with percent grass litter having the highest overall contribution to all models at 31 percent (Table 2-3). Of the ten forest composition variables, four were used for the logistic regression, with canopy cover dominating models (Table 2-3).

The regressions with combined variables all had reasonable concordance (> 75) and over 25 percent decrease in false omission rate and type II error from random, suggesting useful models for predicting occurrence of the three invasive species (Table 2-2). Re-sampling assessment suggested that relative contribution of habitat variables, accuracy for prediction, and p-value were stable for most variables, and models remained significant. There were two variables that were not stable ($p > 0.15$): ammonium for privet and hardwood density for Japanese honeysuckle (Table 2-4). These variables were dropped and the models were rerun.

Chinese lespedeza had a positive relationship with soil magnesium and negative relationships with downed woody debris, midstory cover, and hardwood density. This suggests that Chinese lespedeza is more likely to be found in open or pine areas with higher magnesium levels in the soil and little or no midstory and downed woody debris. There was more than a 50 percent decrease in error from random, suggesting this model is useful in assessing habitat characteristics that are influencing the occurrence of Chinese lespedeza.

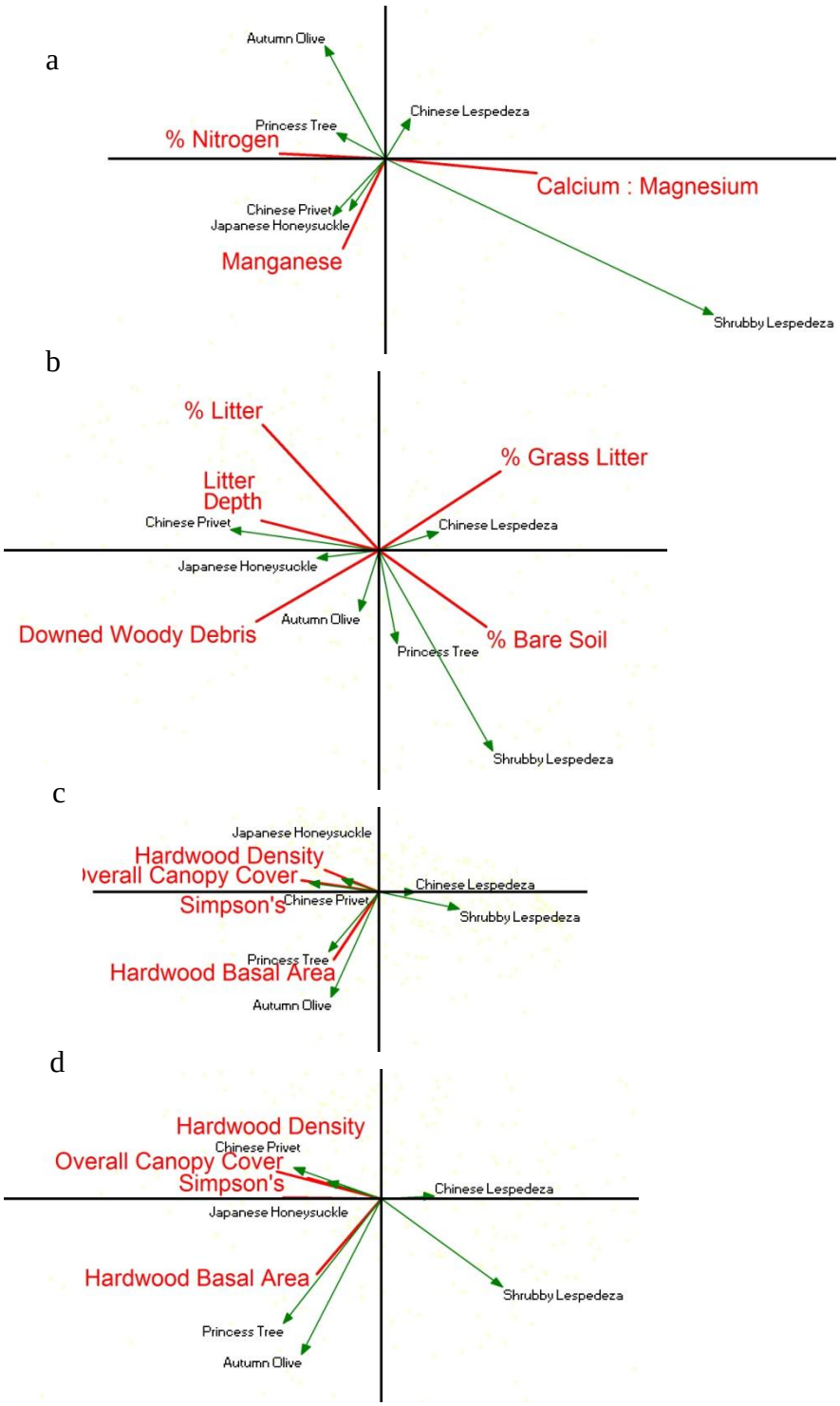

Figure 2-3: Relationship between habitat variables and the invasive community as assessed through Canonical Correspondence Analysis (CCA), a – soil features (axis 1 = 7, axis 2 =5), b – ground (axis 1 =7, axis 2 =2), c – forest structure (axis 1 = 12, axis 2 = 4)), d – all habitat variables combined (axis 1 =13, axis 2 =6), variables $r^2 > 0.30$ are displayed.

Japanese honeysuckle had a positive relationship with canopy cover, soil magnesium, and Simpson's diversity index and a negative relationship with midstory; canopy cover was the most important variable, 41 percent relative importance (Table 2.3). Japanese honeysuckle was found in high canopy cover with little midstory and in areas of high soil magnesium and higher diversity. There was more than a 60 percent decrease in error from random, suggesting this model is useful in assessing habitat characteristics that are influencing occurrence of Japanese honeysuckle (Table 2.2). Chinese privet had one variable, a positive relationship with canopy cover (Table 2.3). There was more than a 50 percent decrease in error from random, suggesting this model is useful in assessing habitat characteristics that are influencing occurrence of privet, even with only one variable (Table 2.3).

Table 2-2: Summary statistics of three invasive species from three logistic regression submodels (soil, ground and forest), combined models and final model (variables that remain stable). MaxSS is the threshold where sensitivity plus specificity is maximized, false omission rate is FN / (FN + TN), type II error is FN / (FN + TP).

		Soil	Ground	Forest	Combined	Final
Chinese Lespedeza	% Concordance	83	75	78	89	89
	Max SS Threshold	0.86	0.68	0.8	0.78	0.78
	Max SS false Omission Rate (%)	68	47	46	34	34
	Max SS Type II (%)	36	11	14	9	9
	Decrease in false omission rate from random (80) (%)	15	41	43	58	58
	Decrease in type II from random (20) (%)	-80	45	30	55	55
Japanese Honeysuckle	% Concordance	76	78	83	85	87
	Max SS Threshold	0.5	0.52	0.56	0.56	0.46
	Max SS false Omission Rate (%)	25	33	30	28	17
	Max SS Type II (%)	14	22	21	20	15
	Decrease in false omission rate from random (58) (%)	57	43	48	52	71
	Decrease in type II from random (42) (%)	67	48	50	52	64
Chinese Privet	% Concordance	64	55	73	77	76
	Max SS Threshold	0.14	0.16	0.2	0.2	0.2
	Max SS false Omission Rate (%)	5	6	7	8	8
	Max SS Type II (%)	7	13	22	25	31
	Decrease in false omission rate from random (19) (%)	74	68	63	58	58
	Decrease in type II from random (81) (%)	91	84	73	69	62

Table 2-3: Summary of significant variables for all three invasive species from three logistic regression submodels (soil, ground and forest), combined models and final model with only variables that remain stable over re-sampling. Variable importance to the model and direction of relationship are given along with the average contribution of each variable to all species.

		Chinese Lespedeza	Japanese Honeysuckle	Chinese Privet	Average Contribution
Soil features	Cation Exchange Capacity		21	66	29
	Magnesium	47	12		20
	Manganese	-25	12		12
	Ammonium	-16	-21	-34	24
	Zinc		-17		6
	% Nitrate		16		5
	Sodium	-12			4
Ground	Downed Woody Debris	-27			9
	Grass Litter		-37	-56	31
	Humus Depth	-31	-8		13
	Shale			-44	15
	% Total Litter		44		15
	Understory	41	12		18
Forest structure	Canopy cover		53	100	51
	Hardwood Density	-77	-9		29
	Midstory	-23	-14		12
	Simpson's		24		8
Variables combined	Canopy cover		45	85	43
	Hardwood Density	-16	-6		7
	Magnesium	32	13		15
	Midstory	-19	-11		10
	Manganese	-33			11
	Ammonium			-15	5
	Simpson's		25		8
Resample assessment based on variables combined	Canopy cover		44	100	48
	Hardwood Density	-16			5
	Magnesium	32	17		16
	Midstory	-19	-18		12
	Manganese	-33			11
	Simpson's		21		7

Table 2-4: Summary of re-sampling of final logistic model for 100 observations run 1000 time. Variable contribution, direction, 99% confidences limit, standard deviations and mean p value of the 1000 re-sampled models is given.

		Contribution			P value
		Mean of re-samples	99% confidence limit	SD	Mean of re-samples
Chinese Lespedeza	Hardwood Density	-13	1.7	6.5	0.11
	Magnesium	38	2.5	9.5	<0.01
	Midstory	-17	2.1	8.2	0.08
	Manganese	-32	2.4	9.0	0.02
Chinese Privet	Canopy cover	81	3.8	14.6	0.04
	Ammonium	-19	3.9	14.9	0.32
Japanese Honeysuckle	Canopy cover	41	2.8	10.6	<0.01
	Hardwood Density	-8	1.6	6.0	0.28
	Magnesium	14	2.6	9.8	0.12
	Midstory	-11	1.9	7.4	0.15
	Simpson's	26	2.3	8.7	0.02

2.5 Discussion

SMCRA mandates that mined land be reclaimed and restored to its original use or a use of higher value. This includes ecosystem functions and services, an integral part of these are the distribution and diversity of the plant species. Restoration assessment often focuses more on the easily measurable restoration of edaphic and hydrological systems. However, these often do not reflect the recovery of the pre-mining biological communities or mitigate landscape, structural, and ecological changes (Burger et al. 2002). Most legislation mandates the evaluation of land reclamation success using readily quantifiable metrics with land assessed after a relatively short time period (Holl & Cairns 2002). This encourages reclamation approaches that address the short-term goals of providing erosion control and minimizing acid mine drainage, but not necessarily the longer-term and more difficult to quantify objective of restoration of ecosystem services. It has been suggested that goals for short-term and long-term recovery of highly disturbed sites may conflict (Holl 2002). Many mine reclamation efforts focus on establishing rapid-growing alien species that control erosion but may slow or prevent the establishment of later-successional, native species (Holl 2002). For example, a general practice creates piles of soil that are then graded to a smooth condition to stabilize the surface and prevent erosion, these sites then are revegetated by hydroseeding with a mixture of herbaceous seeds (mix of grasses and legumes) with fertilizer (Zipper et al. 2011). This can encourage dense herbaceous

vegetation that in turn can negatively affect establishment of native trees and success of planted seedlings (Chaney et al. 1995).

Within our study area, the overall invasive community was most strongly associated with vegetation characteristics such as plant diversity, canopy cover, forest age, and basal area, suggesting that the long-term management of these areas may have the greatest impact on reducing preferential habitat for invasive plants. The majority of these invasive species were in the older, larger, more established forests (15 + years) that had higher tree diversity and where the invasive species would have had more time to establish. The managed monoculture pine plantations and open areas were less likely to have multiple invasive plants.

Forest characteristics dominated both the CCA and regression models. Canopy cover, basal area, age, Simpson's index, midstory percent cover, and hardwood density were the most useful environmental variables. Four of the species are strongly associated within the community analysis, Chinese privet, autumn olive, princesstree, and Japanese honeysuckle, suggesting similar habitat preferences.

Species-by-species logistic regression models for the three that had sufficient data revealed some differences. Chinese lespedeza has been planted since 1970 as part of reclamations; this still continues today (pers com Dr. Randall Johnson, Director, ASMC). It is very prevalent throughout the SHR, having been widely planted and then dispersed. Its high tolerance for a wide variety of habitats (Miller et al. 2010) has made it a pervasive invader in the area. It forms thick clusters that have spread over large areas and may ultimately prevent forest regeneration (Miller et al. 2010). In this study, Chinese lespedeza was more likely to be found in open or pine areas with higher magnesium levels in the soil and little or no midstory and downed woody debris. The model had a high false omission rate, suggesting there are other reasons for Chinese lespedeza occurrence than the attributes measured. One of the potential confounding factors is the active planting of this species. For the management of this species, increased canopy cover with a diverse forest structure seems to be the best long-term approach, but the biggest contribution to management of this species would be elimination from seeding material.

Japanese honeysuckle has been widely planted for deer and cattle forage (Dickson et al. 1978; Patterson 1976) and is now considered naturalized in upland and lowland forests as well as in forest-edge habitats (Patterson 1976; Yates et al. 2004). It is not as detrimental as some of the other alien species, but it has been shown to impact even-aged pine regeneration when at very high densities. In this study, Japanese honeysuckle was found in areas with high canopy cover

with little midstory, low density of hardwoods, and in areas of high soil magnesium and higher diversity.

Of the three species considered, Chinese privet might be the most detrimental. It is considered the second most abundant invasive plant in the South and is most prevalent in the understory of bottomland hardwood forests (Merriam & Feil 2002). It can form dense stands to the exclusion of most native plants and replacement regeneration, impacting the abundance of specialist birds and diversity of native plants and bees (Wilcox & Beck 2007). In this study, Chinese privet was found in high canopy cover areas, however the model was not strong, suggesting there are other factors influencing its distribution.

The influence of planting alien, invasive species in this area is likely the major driver of the high diversity of invasive plants, with three of the six dominant species being planted. Adjusting the reclamation plantings to native species would aid in resolving this. In terms of the impact these species are having on the reclamation and productivity of the land, further study needs to be undertaken. Of the three most dominant species, one is planted and another is ubiquitous throughout the region at low densities. The third species, privet, is of most concern. Overall, it appears that the initial reclamation efforts, apart from the planting of invasive species, are not the major driver impacting the alien, invasive species composition of the reclaimed, now forested mine sites.

2.6 Literature Cited

Alpert, P., Bone, E. & Holzapfel, C. (2000) Invasiveness, invasibility and the role of environmental stress in the spread of non-native plants. Perspectives in Plant Ecology, Evolution and Systematics, 3, 52-66.

Bradshaw, A.D. (1984) Technology lecture: Land restoration: Now and in the future. Proceedings of the Royal Society of London. Series B, Biological Sciences (1934-1990), 223, 1-23.

Brown, J.A. (2003) Designing an efficient adaptive cluster sample. Environmental and Ecological Statistics, 10, 95-105.

Burger, J.A., Scott, D. & Mitchem, D. (2002) Field assessment of mine soil quality for establishing hardwoods in the Appalachians. Reclamation with a purpose. (eds R. Barnhisel & M. Collins), pp. 226-240. 19th Annual Meeting, American Society of Mining and Reclamation, Lexington, KY.

Chaney, W.R., Pope, P.E. & Byrnes, W.R. (1995) Tree survival and growth on land reclaimed in accord with Public Law 95-87. Journal of Environmental Quality, 24, 630-634.

Davison, V.E. (1945) Wildlife values of the lespedezas. Journal of Wildlife Management, 9, 1-9.

Dickson, J.G., Segelquist, C.A. & Rogers, M.J. (1978) Establishment of Japanese honeysuckle in the Ozark Mountains. Proceedings of the Southeastern Association of Fish and Wildlife Agencies. pp. 242-245.

Donovan, M.L., Mayhew, S.L., Warren, B.E. & Stephens, V.E. (2007) An evaluation of chemical treatment and burning on the control of autumn olive (*Elaeagnus umbellate*). Report No. 3471. Michigan Department of Natural Resources, Wildlife Division. Lansing, MI

Dyess, J.G., Causey, M.K., Striblin, H.L. & Lockaby, B.G. (1994) Effects of fertilization on production and quality of Japanese honeysuckle. Southern Journal of Applied Forestry, 18, 68-71.

D'Antonio, C.M., Dudley, T.L. & Mack, M.C. (1999) Disturbance and biological invasions: Direct effects and feedbacks. Ecosystems of disturbed ground. (ed L. Walker), pp. 413-452. Elsevier, Amsterdam.

Evans, C. W., Moorehead, D.J., Bargeron, C. T., & Douce, G. K. (2006) Invasive Plant Responses to Silvicultural Practices in the South. The University of Georgia Bugwood Network, Tifton, GA. BW-2006-03. 52pp

Graham, E.H. (1941) Legumes for erosion control and wildlife. USDA Misc. Publ. No. 412, Washington D.C. 153pp

Greenburg, C.H. & Walter, S.T. (2010) Fleshy fruit removal and nutritional composition of winter-fruiting plants: a comparison of non-native invasive and native species. Natural Areas Journal, 30, 312-321.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. & Tatham, R.L. (2006) Multivariate Data Analysis Sixth Edition Pearson Education. New Jersey.

Hanula, J.L., Horn, S. & Taylor, J.W. (2009) Chinese privet (*Ligustrum sinense*) removal and its effect on native plant communities of riparian forests. Invasive Plant Science and Management, 2, 292-300.

Haugen, A.O. & Fitch, F.W.J. (1955) Seasonal availability of certain bush lespedeza and partridge pea seed as determined from ground samples. Journal of Wildlife Management, 19, 297-301.

Heikkinen, R.K. (1996) Predicting patterns of vascular plant species richness with composite variables: A meso-scale study in Finnish Lapland. Plant Ecology, 126, 151-165.

Hill, M.O. (1991) Patterns of species distribution in Britain elucidated by canonical correspondence analysis. Journal of Biogeography, 18, 247-255.

Holl, K.D. (2002) Long-term vegetation recovery on reclaimed coal surface mines in the eastern USA. Applied Ecology, 39, 960-970.

Holl, K.D. & Cairns, J.J. (2002) Monitoring ecological restoration. Handbook of Ecological Restoration. pp. 411-432. Cambridge University Press, Cambridge, UK.

Hosmer, D.W. & Lemeshow, S. (2000) Applied Logistic Regression. Wiley-Interscience, New York.

Johnson, J.E., Mitchem, D.O. & Kreh, R.E. (2003) Establishing royal paulownia on the Virginia Piedmont. New Forests, 25, 11–23.

Kriticos, D.J., Sutherst, R.W., Brown, J.R., Adkins, S.W. & Maywald, G.F. (2003) Climate change and the potential distribution of an invasive alien plant: *Acacia nilotica* ssp. *indica* in Australia. Journal of Applied Ecology, 40, 111-124.

Langdon, K.R. & Johnson, K.D. (1994) Additional notes on invasiveness of *Paulownia tomentosa* in natural areas. Natural Areas Journal, 14, 139-140.

Leatherman, A.D. (1955) Ecological life-history of *Lonicera japonica* Thunb. Dissertation. University of Tennessee, Knoxville, TN, USA.

Lindsay, W.L. & Norvell. (1978) Development of a DTPA soil test for zinc, iron, manganese, and copper. Soil Science Society of America Journal, 42, 421-428.

Lodge, D.M. (1993) Biological invasions: Lessons from ecology. Trends in Ecology & Evolution, 8, 133-136.

Longbrake, A.C.W. (2001) Ecology and invasive potential of *Paulownia tomentosa* (Scrophulariaceae) in hardwood forest landscape. PhD, Ohio University, Plant Biology, 174.

Longbrake, A.C. & McCarthy, B.C. (2001) Biomass allocation and resprouting ability of princess tree (*Paulownia tomentosa*: Scrophulariaceae) across a light gradient. The American Midland Naturalist, 146, 388-403.

Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology, 38, 921-931.

Mantel, N. and Valand, R. S. (1970) A technique of nonparametric multivariate analysis. Biometrics 26: 547-558

McGrath, D. A., Evans, J.P., Smith, C.K., Haskell, D.G., Pelkey, N.W., Gottfried, R.R., Brockett, C.D., Lane, M.D. & Williams, E.D. (2004) Mapping land-use change and monitoring the impacts of hardwood-to-pine conversion on the Southern Cumberland Plateau in Tennessee. Earth Interactions, 8, 1.

McSweeney, K. & Jansen, I.L. (1984) Soil structure and associated rooting behaviour in mine soils. Soil Science Society of America Journal, 48, 607-612.

Mehlich, A. (1984) Mehlich 3 soil test extractant: A modification of the Mehlich 2 Extractant. Communications in Soil Science & Plant Analysis. 15, 1409-1416.

Merriam, R.W. & Feil, E. (2002) The potential impact of an introduced shrub on native plant diversity and forest regeneration. Biological Invasions, 4, 369-373.

Miller, J.H., Chambliss, E.B. & Loewenstein, N.J. (2010) A field guide for the identification of invasive plants in southern forests. Gen. Tech. Rep. SRS–119. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 126 p.

Miller, J.H., Lemke, D. & Coulston, J. (2012) The Southern Forest Futures Project, Chapter 15. The Invasion of Southern Forests by Nonnative Plants: Current and Future Occupation with Impacts, Management Strategies, and Mitigation Approaches. In Press.

Negley, T.L. & Eshleman, K.N. (2006) Comparison of storm flow responses of surface-mined and forested watersheds in the Appalachian Mountains, USA. Hydrological Processes, 20, 3467-3483.

Nilsson, L. & Belyaev, Y.K. (1998) Application of resampling to exponential and logistic regression. Research Report 1, Department of Mathematical Statistics, Umea University, Sweden.

Ohwi, J. (1965) Flora of Japan (FG Meyer and EH Walker, Eds.). Smithsonian Institution, Washington, DC, USA.

Oommen, T., Baise, L.G. & Vogel, R.M. (2010) Sampling Bias and Class Imbalance in Maximum-likelihood Logistic Regression. Mathematical Geosciences, 43, 99-120.

Patterson, D. (1976) The history and distribution of five exotic weeds in North Carolina. Southern Appalachian Botanical Society, 41, 177-180.

Ricciardi, A. (2007) Are modern biological invasions an unprecedented form of global change? Conservation Biology, 21, 329-36.

Ricketts, T.H., Dinerstein, E., Olson, D.M., Loucks, C.J. & al., et. (1999) Terrestrial ecoregions of North America: a conservation assessment. Island Press, Washington, D.C.

Ripley, E.A., Redman, R.E. & Crowder, A.A. (1996) Environmental effects of mining. St. Lucie Press, Boca Raton, Florida, USA.

Sasek, T.W. & Strain, B.R. (1990) Implications of atmospheric CO^2 enrichment and climatic change for the geographical distribution of two introduced vines in the U.S.A. Climatic Change, 16, 31-51.

Sather, N. & Eckardt, N. (1987) Element Stewardship Abstract for *Elaeagnus umbellata* (Autumn olive). Arlington, VA.

Schierenbeck, K. (2004) Japanese honeysuckle (*Lonicera japonica*) as an invasive species; history, ecology, and context. Critical Reviews in Plant Sciences, 23, 391-400.

Shannon, C.E. & Weaver, W. (1949) The mathematical theory of communication. Univ. Illinois Press, Urbana.

Sharp, W.C. (1977) Conservation plants for the Northeast. USDA Soil Conservation Service, Program Aid No. 1154. U.S. Government Printing Office, Washington, D.C. 40pp

Shelton, M.G. & Cain, M.D. (2002) Potential carry-over of seeds from 11 common shrub and vine competitors of loblolly and shortleaf pines. Canadian Journal of Forest Research, 32, 412-419.

Shukla, M.K., Lal, R. & Ebinger, M.H. (2005) Physical and chemical properties of a mine spoil eight years after reclamation in northeastern Ohio. Soil Science Society of America Journal, 69, 1288-1297.

Simberloff, D., Schmitz, D. & Brown, T. (1997) Strangers in paradise: impact and management of nonindigenous species in Florida. Island Press, Washington, DC, USA.

Simpson, E.H. (1949) Measurement of diversity. Nature, 163, 688.

Skelly and Loy Engineers (1979) A Compliance Manual – Methods for meeting OSM requirements. McGraw Hill

Skousen, J.G., Johnson, C.D. & Garbutt, K. (1994) Natural revegetation of 15 abandoned mine land sites in West Virginia. Journal of Environmental Quality, 23, 1224-1230.

Smalley, G.W. (1979) Classification and evaluation of forest sites on the Southern Cumberland Plateau. Gen. Tech. Rep. SO-23. U.S. Department of Agriculture, Forest Service. Southern Forest Experiment Station, New Orleans, LA, 59 p.

Stevens, D.L. & Olsen, A.R. (2004) Spatially balanced sampling of natural resources. Journal of the American Statistical Association, 99, 262-278.

Sun, Q.B., Shen, R.F., Zhao, X.Q., Chen, R.F. & Dong, X.Y. (2008) Phosphorus enhances Al resistance in Al-resistant *Lespedeza bicolor* but not in Al-sensitive *L. cuneata* under relatively high Al stress. Annals of Botany, 102, 795-804.

Travis, J. & Wilterding, J. (2005) Assessment of autumn olive (*Elaeagnus umbellate*) population at Pierce Cedar Creek. Olivet College Olivet, MI.

USDA (2011) PLANTS National Plants Database. National Plant Data Center, Baton Rouge. http://plants.usda.gov.

USDA Forest Service (1998) Field Instructions for Southern Forest Inventory. Southern Research Station.

Vitousek, P.M., D'Antonio, C.M., Loope, L.L., Rejmanek, M. & Westbrooks, R. (1997) Introduced species: A significant component of human-caused global change. New Zealand Journal of Ecology, 21, 1-16.

Wear, D.N. & Greis, J.G. (2002) Southern forest resource assessment: summary of findings. Journal of Forestry, 100, 6–14.

Webster, C.R., Jenkins, M.A.M.A. & Jose, S. (2006) Woody invaders and the challenges they pose to forest ecosystems in the eastern United States. Journal of Forestry, 104, 366-379.

Wilcox, J. & Beck, C.W. (2007) Effects of *Ligustrum sinense* lour. (Chinese privet) on abundance and diversity of songbirds and native plants in a southeastern nature preserve. Southeastern Naturalist, 6, 535-550.

Williams, C.E. (1993) The exotic empress tree, *Paulownia tomentosa*: an invasive pest of forests? Natural Areas Journal, 13, 221-222.

Yates, E.D., Levia, D.F. & Williams, C.L. (2004) Recruitment of three non-native invasive plants into a fragmented forest in southern Illinois. Forest Ecology and Management, 190, 119-130.

Zeleznik, J.D. & Skousen, J.G. (1996) Survival of three tree species on old reclaimed surface mines in Ohio. Journal of Environmental Quality, 25, 1429-1435.

Zipper, C.E., Burger, J.A., McGrath, J., Rodrigue, J.A. & Holtzman, G.I. (2011) Forest Restoration Potential of Coal-Mined Lands in the Eastern United States. Journal of Environmental Quality, 40, 1567-1577.

CHAPTER 3.

Geospatial Assessment of Non-native Vegetation: application of geospatial models to predicting alien plants on reclaimed mines in the Shale Hills region.

Collaborators: *C. J. Schweitzer, W. Tadesse, and Y. Wang*

3.1 Abstract

Throughout the world, the invasion of alien plants is an increasing threat to native biodiversity. Invasion is especially prevalent in areas affected by land transformation and anthropogenic disturbance. Surface mines are a major disturbance, and this may promote the establishment and development of invasive plant communities. Using the Shale Hills Region of Alabama as a case study, we examined the landscape characteristics, as measured through remotely sensed and other geospatial datasets, to predict the probability of occurrence of six invasive plant species. Mine lands were surveyed for all species defined by the United States Forest Service as invasive to the forest of the southeastern region. This paper focuses on species with more than 50 occurrences across the sampled landscape: Chinese lespedeza (*Lespedeza cuneata*), Japanese honeysuckle (*Lonicera japonica*), Chinese privet (*Ligustrum sinense*), autumn olive (*Elaeagnus umbellata*), princesstree (*Paulownia tomentosa*) and sawtooth oak (*Quercus accitimus*). Three of the species have and continue to be planted as part of reclamation activities: Chinese lespedeza, autumn olive and sawtooth oak. Invasive species occurrence data were assessed using logistic regression and maximum entropy modelling, integrated with geospatial data. We used an area under the curve value for the receiver operator characteristic of greater than 0.75 and decrease omission rate of more than 0.25 as defining a good model. The distance to forest had the highest overall contribution (19%) to the models, with three other variables have over 10 percent relative importance, distance to roads, Normalized Difference Vegetation Index in 1987 and 2011. Species models were then applied to the mined landscape to assess the probable prevalence of each species across the landscape. Japanese honeysuckle had the highest probable prevalence at 48 percent, with princesstree having the lowest, at less than 1

percent. Overall 33 percent of the landscape is predicted to have no invasive plants, with 47 percent predicted to have one, 17 percent two, and 3 percent to have three or more. Japanese honeysuckle, lespedeza, privet and autumn olive showed much higher occurrence on the reclaimed sites than that from across the broader region. We found that geospatial modelling of these invasive plants, at this scale, was useful and does offer potential for management, both in terms of identifying habitat types most at risk and identifying areas needing management attention.

3.2 Introduction

Land transformation and anthropogenic disturbance often facilitate the establishment and development of invasive plant community. Surface mining is a major is form of disturbance and has affected over 2.4 million hectares of terrestrial habitat in the United States since 1930 (Zeleznik & Skousen 1996). It is essential that the impact of this mining be mediated to reduce its impact on the surrounding environment. Throughout the world, alien plants are becoming an increasing threat to native biodiversity and ecosystem functions (Ricciardi 2007; Vitousek et al. 1997). Invasion often occurs after disturbance (Alston & Richardson 2006), and in the Shale Hills Region (SHR) there are over 300 mines that have been permitted since 1977. These are major sources of disturbance to the landscape, and thus may be a major source of invasive plants to the surrounding forest.

Since the introduction of the Surface Mine Reclamation Act (SMCRA) in 1977, much of the land transformed by surface mining has been reclaimed, with efforts aimed at improving the quality of the land by restoring some pre-disturbance vegetation and functions (Bradshaw 1984). The SMCRA mandates that mined land be reclaimed and restored to its original use or a use of higher value. However, surface mine reclamation efforts rarely result in ecosystems that simulate pre-mined characteristics. In the past the focus on edaphic and hydrologic systems has not been conducive to achieving goals related to the recovery of the pre-mining biological communities or mitigation of landscape structure and ecological change (Burger et al. 2002). However, SMCRA states that mining operations shall establish "*a diverse, effective, and permanent vegetative cover of the same seasonal variety and native to the area and capable of self-regeneration and plant succession …, unless use of introduced species is necessary to achieve the stated post-mining land use*", and industry goals of reclamation are changing (Zipper et al. 2011). Still, most legislation mandates evaluation of land reclamation success after a relatively short time period, if

at all (Holl & Cairns 2002). This encourages reclamation approaches that address the short-term concerns of providing erosion control and minimizing acid mine drainage, but not long-term concerns of restoration of ecosystem services. For example, the success of coal surface mine reclamation efforts in the southeastern United States is usually evaluated within five years (McElfish & Beier 1990). Such practices have resulted in large patches of grassland not typical of the native landscape (Hardt & Forman 1989). It has been suggested that goals for short-term and long-term recovery of highly disturbed sites may conflict (Holl 2002). Planting aggressive alien ground cover species to minimize short-term erosion may slow long-term recovery. Recent work has shown the cost of historically used restoration methods are comparable with those that develop a more diverse and ecologically sound community (Zipper et al. 2011). However, the legacy of alien plants remains, even though there is a transition to a more ecologically stable restoration approach. Thus understanding the distribution across the landscape is an important component of invasive plant management and for evaluating the incidence of, and the potential for invasion (Davis et al. 2000). Ground-based assessments are costly but provide essential information. Techniques such as remote sensing offer significant opportunities for providing information on invasions of alien species and can be used to assess the broader vegetation characteristics. Remote sensing affords the opportunity to map an entire region and assessing invasive species distribution in less accessible areas. Imaging techniques can also offer financial advantages over field-based approaches, and accurately delineating the spatial distribution of invasives and combining maps with other environmental and anthropogenic data layers can provide the basis for predicting expansion into surrounding areas (Underwood et al. 2007)

There have been two approaches used in imaging, high spatial but low spectral resolution (e.g. black and white or colour infrared aerial photographs) or low spatial with high spectral resolution (e.g. Landsat). There are costs and benefits to both. High spatial, but low spectral resolution often requires high processing time and works only on species that possess visually detectable unique characteristics, making it useful for small areas. Low spatial but high spectral resolution give more opportunity for automated image processing, access to recent historical data for time series analyses, and large spatial coverage, however low spatial resolutions often means invasive species populations can often only be detected once they are dense and widespread (Underwood et al. 2007).

In this study we assess the uses of remotely sensed and other geospatial datasets, at multiple resolutions, to develop non-field based models for assessment of non-native invasive plants on

reclaimed mines. Both traditional statistics and machine learning techniques are used to model invasive probabilities across the mined landscape of the SHR.

3.3 Methods

3.3.1 Study Area

Our study was conducted in the Shale Hills Region (SHR) of the southern Cumberland Plateau of the south-eastern United States (Figure 2-2), as described in Chapter 2. Our target area included surface mines permitted after 1983 that were closed before 2006, thus had time to be reclaimed and for vegetation to re-establish. The final phase of restoration is planting of the permanent vegetation, mines considered in this study were planted at a rate of 500 – 700 pines per acre, with 450 tree per acre considered successful (Skelly and Loy 1979).

3.3.2 Species of Interest

Mine lands were surveyed for all species defined by the United States Forest Service (USFS) as invasive to the forest of the southeastern region (Miller et al. 2010). This chapter focuses on species with more than 50 occurrences across the sampled landscape, and included six species: Chinese lespedeza (*Lespedeza cuneata*) (n=2475, 53%), Japanese honeysuckle (*Lonicera japonica*) (n=1403, 30%), Chinese privet (*Ligustrum sinense*) (n=238, 1%), autumn olive (*Elaeagnus umbellata*) (n=436, 2%), princesstree (*Paulownia tomentosa*) (n=126 0.5%) and sawtooth oak (*Quercus accitimus*) (n=62 <0.5%). Three of the species have been and continue to be, planted as part of reclamation activities; Chinese lespedeza, autumn olive and sawtooth oak. Descriptions of species are given in Chapter 2, apart from sawtooth oak, which was not found on the vegetation plots.

3.3.2.1 Sawtooth Oak (*Quercus accitimus*)

Introduced from Asia in 1862 as an ornamental, sawtooth oak is a member of the white oak family (Short 1976). It is native to Japan, Korea, China, and areas of the Himalayans (Huntley & Hopkins 1979). Sawtooth oak has rapid growth and prolific mast production at an early age and as such may be useful for wildlife. The sawtooth oak reaches heights of 20 m at maturity, and mast production may begin as early as six years of age, with regular annual production by ten years of age (Huntley & Hopkins 1979). Although sawtooth oak possesses many favourable

traits, some studies have shown that it is not as hardy as some of the native oaks and may not be as long lived (Huntley & Hopkins 1979). It has been planted as part of reclamation in this area since the 1970s (pers com Dr. Randall Johnson, Director, Alabama Surface Mining Commission [ASMC]).

3.3.3 Site Selection

Sites were selected using stratified spatial balanced sampling, a Generalized Random Tessellation Stratified (GRTS) design (Stevens & Olsen 2004). Sixty groups of plots were located across the 1983 – 2006 reclaimed mined landscape of the SHR. Thirty-six groups of sites were visited Sampling was carried out in the spring of 2011. Sampling was stratified by mine age: >20 years, 10 – 20 years and <10 years. Occurrence of invasive plants was assessed for every 6 m of the transect. Forbs and vines were only assessed on the main transect, while trees and shrubs were assessed on the main transect, as well as 3 – 9 m and 9 – 15 m on either side of the main transect. Vegetation sampling for herbaceous plants, forbs and vines was carried out at 129 plots within each site, and at 645 plots for shrubs and trees to give a total of 4,644 plots assessed for herbs, vines and forbs and 23,220 plots for trees and shrubs. Plots were arranged in figure eight transects (crossing at right angles in the middle of the transect) covering 1.6 km (200 m on each side of the figure eight).

3.3.4 Geospatial Data

Mine boundaries were obtained from Alabama Surface Mining Commission and verified with aerial photos. The time since mine closure was determined by the permit release or forfeit date (forfeited permits often did not undergo full reclamation, with natural re-establishment of plant communities), and grouped into three age classes (1983 – 1990, 1991 – 2000, and 2001 – 2006). Environmental and topographic variables were represented by slope, aspect (northness), solar radiation, curvature, and distance from water. These variables were selected based on their biological significance in other studies (Bartuszevige et al. 2006; Gutirrez et al. 2005; Lemke et al. 2011; Lockwood et al. 2007). They were predominately derived from a 10 m digital elevation model (DEM) (Gesch et al. 2002). The DEM was used to generate slope (degrees), aspect (degrees), solar radiation (Wh m^{-2}) and curvature using ArcGIS (ESRI 20010), Spatial Analyst Tools. Aspect was transformed into a linear north–south gradient (northness) by performing cosine transformation (Guisan et al. 1999). Solar radiation was calculated as the annual watt-

hours per square metre given no cloud cover. Curvature is a measure of shape of the landscape, whether it is flat, convex, or concave. In ArcGIS "curvature" assesses surrounding cells to calculate a curvature, with increasing positive scores representing increasing concavity (ESRI 2010).

Streams and water bodies may affect the distribution and establishment of plant species by influencing moisture availability. Riparian areas have been shown to contain more alien plant species than nearby upland areas (Stohlgren et al. 2002). Therefore distance from stream was included in the model. Considerable landscape alteration (due to mining activities) has occurred so this information was digitized from 2009 aerial photography and not taken from available geospatial datasets. Climate and elevation data were not integrated into these models. Since only a few counties were sampled, variation is climate and elevation is not likely to be significant.

A number of anthropogenic variables were integrated into this study including land cover and roads. Public road files are available from the Census Bureau and Alabama Department of Transport, but due to the large numbers of access roads additional roads were digitized from 2009 aerial photos. Land cover data were taken from the National Land Cover Data (NLCD) available for 1992, 2001 and 2006 (Fry et al. 2011; Homer et al. 2004; Vogelmann et al. 2001). Land cover characteristics that were derived for the above dataset included percent forest within 100 m of plot for each year. Open area was digitized from aerial photos of 2009. Distance to forest edge was also estimated using this dataset.

Colour infrared imagery (CIR) for 2009, Landsat Thematic Mapper imagery of 1987, 1991, 1998, 2004 and 2011 were used to derive Normalized Difference Vegetation Index (NDVI) (Rouse et al. 1974). The NDVI is a simple indicator used to analyse remote sensing measurements to assess whether the target being observed contains live green vegetation or not. Images were pre-processed by absolutely calibrating the most recent image (reference image), and then normalizing the older historical images.

3.3.5 Data Analysis

Correlations between each of the environmental and anthropogenic variables were assessed in SAS (SAS Institute 2009) to identify less correlated ($r^2 < 0.50$) variables for further analysis (Table A9-5). Invasive species occurrence data were assessed using logistic regression and MaxEnt modelling approaches. Logistic regression is a generalized linear model that is used to investigate the relationship between a categorical outcome and a set of explanatory variables or

for prediction of the probability of occurrence of an event by fitting data to a linearization of the logistic curve, using the absence/presence data. It makes use of several predictor variables that may be either numerical or categorical. Logistic regression makes no assumptions about the distribution of the independent variables. Logistic regression was conducted in SAS. MaxEnt (Phillips et al. 2006) is based on a maximum entropy probability distribution, the distribution whose entropy is at least as great as that of all other members of a specified class of distributions. MaxEnt estimates the probability distribution that is most spread out subject to constraints such as the environmental characteristics at known locations of the species. The MaxEnt model only uses occurrence data. To assess models, data were split spatially with 30% assigned to a test set and 70% to a training set, with one hundred replications run for the MaxEnt and logistic regression models to obtain the average contribution, Area Under the Curve (AUC) and type II error. For logistic regression, sample data were re-sampled to give at least a 20% occurrence for each species (Oommen et al. 2010). For descriptive purposes variable importance and direction of variables were tabulated. For logistic regression variable importance was calculated using the Wald chi squared statistic, dropping the intercept Wald chi square and standardizing the remainder to 100. Accuracy of prediction was assessed using AUC and type II error (FN/(FN+TP). Type II error was assessed based on a threshold value determined by maximizing specificity plus sensitivity (Manel et al. 2001). Due to variation in species occurrence across the study area, benchmark type II errors were defined as if data were randomly assigned, and decrease of more than 25% was considered a useful model (Manel et al. 2001;Hair et al. 2006). We used the following classes of AUC to assess model performance: 0.50 – 0.75 = fair, 0.75 – 0.92 = good, 0.92 – 0.97 = very good, and 0.97 – 1.00 = excellent (Hosmer & Lemeshow 2000). Models were then applied to the geospatial data to give potential distribution of each species across the area. Maps were generated by reclassifying the continuous output to binary using the maximized specificity and sensitivity threshold. For each species the logistic regression and MaxEnt maps were combined to give estimates of the proportion of the landscape that had low potential (not predicted by either model), moderate potential (predicted by one model) and high potential (predicted by both models). The spatial correlations between models for each species and between species were measured through ArcGIS band collection statistics correlation matrix. These maps were then combined to give an estimate of invasive species diversity across the landscape.

3.4 Results

Overall elevation ranged from 103 m to 230 m, with an average slope of 9 percent. The average distance to a road (including service roads) was 92 m and was 240 m to water, and the surrounding area (100 m radius) of any point was on average 42 percent forested (Table 3-1). Pearson's correlation was used to remove the highly correlated variables of open area, proportion of forest within a 100 m area, NDVI 1991 and NDVI 2004 resulting in 13 variables for model use (Table A9-5).

Table 3-1: Summary of geospatial variables measured at each sampling site.

Code	Variable	mean	SD	Range	Unit	Source
Age	Years since bond release or forfeit	1995	6	1983 to 2006	years	Alabama Surface Mining Commission
Northness	transformed into a linear north–south gradient, cosine transformation	0.03	0.71	-1 to 1		USGS 10 m DEM
Curvature	shape of the landscape, whether it is flat, convex, or concave (increasing positive scores representing increasing concavity)	0.04	1.46	-14.8 to 16.5		USGS 10 m DEM
Slope	Slope	8.9	6.6	0.3 to 55	degrees	USGS 10 m DEM
Solar	Solar radiation given no cloud cover	246	10	131 to 254	Wh/m^2	USGS 10 m DEM
Dist River	Distance to water	240	132	0 to 660	m	2009 aerial photo
Open100	Proportion of open land within 100 m of plot	0.39	0.34	0 to 1		2009 aerial photo
Dist Forest	Distance to forest	73.3	88.1	0 to 360	m	2009 aerial photo
Dist Roads	Distance to roads	92.2	71.6	0 to 352	m	2009 aerial photo
Forest06	Proportion of forest within 100 m of plot in 2006	0.42	0.36	0 to 1		USGS LULC
Forest00	Proportion of forest within 100 m of plot in 2001	0.46	0.36	0 to 1		USGS LULC
Forest90	Proportion of forest within 100 m of plot in 1992	0.78	0.29	0 to 1		USGS LULC
NDVI2009	NDVI in 2009	0.19	0.05	-0.05 to 0.31		2009 aerial photo
NDVI1987	NDVI in 1987	0.43	0.25	-0.12 to 0.74		Landsat
NDVI1991	NDVI in 1991	0.34	0.21	-0.05 to 0.71		Landsat
NDVI1998	NDVI in 1998	0.23	0.11	-0.05 to 0.6		Landsat
NDVI2004	NDVI in 2004	0.41	0.17	-0.03 to 0.73		Landsat
NDVI2011	NDVI in 2011	0.43	0.11	0 to 0.66		Landsat

A good model was defined as one that had an AUC of greater than 0.75 and decrease type II error of more than 0.25, and nine good models were developed. Lespedeza was the only species that did not have any good models. Each of the other five species had similar AUC and decrease

in type II errors between the MaxEnt and logistic regression models. The greatest discrepancy was for sawtooth oak where MaxEnt had an AUC 0.09 higher than logistic regression and 0.15 greater decrease in type II error than logistic regression. The distance to forest had the highest overall contribution (19%) to the models, with distance to roads, NDVI in 1987 and 2011 each have over 10 percent relative importance. All dominant variables (over 10% relative importance to any model) had the same direction of relationship with the different model types.

At an individual species level, Chinese lespedeza had the weakest models with logistic regression test AUC of only 0.70 and MaxEnt test AUC of 0.69 and decrease in type II errors of 0.45 and 0.36, respectively (Table 3.2). This suggests that the model may be useful but not strong. The occurrence of lespedeza is predicted to increase in new mines with characteristics including greater distance from forest, closer distance to roads, less forest in 1990, a lower NDVI in 1998 and 2011. Slope, northness, NDVI 2009 and curvature made very little contribution to determining occurrence of lespedeza.

Japanese honeysuckle had the second weakest model with logistic regression test AUC of 0.75 and MaxEnt test AUC 0.73, however these models had a high decrease in test type II errors at 0.70 for logistic regression and 0.66 for MaxEnt (Table 3.2). Two variables, distance to forest and NDVI in 2011, dominated both models, with relative importance of over 60 percent of each model's. Japanese honeysuckle had a higher chance of occurrence when associated close to or within a forest and with a higher NDVI in 2011.

Privet had reasonable models, with AUC of 0.79 and 0.83, and decreases in type II errors of 0.54 and 0.58 (Table 3.2). The models agreed that a higher chance of privet occurrence was associated with proximity to forest(s), proximity of river(s) and age (older mines as having higher chance of occurrence).

Autumn olive had strong models with AUC of 0.82 and 0.88 and change in type II errors of 0.85 and 0.78 (Table 3.2). All variables had similar relationships across models with autumn olive more likely to occur closer to, or within forest, closer to rivers but further from roads and with low NDVI1987.

Princesstree had strong models with AUC of 0.97 and 0.96 and change in type II errors of 0.91 and 0.93 (Table 3.2). Variables had similar relationships between models with NDVI in 1987 dominating both models (negatively), and also higher probability of occurrence in new mines with a high NDVI in 2009.

Sawtooth oak had reasonably good models with AUC's of 0.83 and 0.92 and change in type II errors of 0.60 and 0.75 (Table 3.2). Overall, the models suggested sawtooth oak is found close to, or within forests, and further from roads.

Table 3-2: Summary statistics of the logistic regression and MaxEnt models from 100 resample's for six dominant invasive plant species, variable contribution and direction of relationship(- negative, + positive, ∩ or U for binomial relationship) are given along with the average contribution of each variables to all species.

	Autumn Olive		Chinese Lespedeza		Privet		Japanese Honeysuckle		Princesstree		Sawtooth Oak	
	M	L	M	L	M	L	M	L	M	L	M	L
Test AUC	0.88	0.82	0.69	0.70	0.83	0.79	0.73	0.75	0.96	0.97	0.92	0.83
MaxSS Threshold	0.31	-1.83	0.46	-0.09	0.32	-1.26	0.42	-0.88	0.11	-1.30	0.27	-1.06
MaxSS type II error	0.17	0.12	0.30	0.26	0.34	0.37	0.24	0.21	0.06	0.07	0.20	0.32
Random type II error	0.80	0.80	0.47	0.47	0.80	0.80	0.70	0.70	0.80	0.80	0.80	0.80
Decrease in type II error from random	0.78	0.85	0.36	0.45	0.58	0.54	0.66	0.70	0.93	0.91	0.75	0.60
Mine age	2	6	5	22	-10	-15	-2	-2	17	12	0	1
Curvature	1	1	0	0	∩1	0	∩1	0	0	1	∩1	0
Dist Forest	-17	-13	7	12	-21	-41	-27	-42	-2	2	-15	-34
Dist River	-12	0	8	4	-21	-24	U3	-3	0	-1	∩1	3
Dist Roads	7	14	U4	-5	-8	-1	U5	-5	∩2	7	30	36
Fores06	13	0	-5	2	∩9	-8	3	-11	4	1	∩18	-4
Forest92	-3	-9	-10	-11	-1	0	2	3	-1	0	-1	-9
NDVI1987	-35	-51	U2	-7	∩10	1	4	0	-60	-45	∩25	-3
NDVI1998	∩2	2	∩26	-28	∩1	0	11	7	0	-3	0	-1
NDVI2009	1	1	∩2	-1	-1	1	1	4	9	12	3	4
NDVI2011	5	2	∩29	-7	∩12	-3	38	22	3	7	U1	-3
Northness	1	1	1	0	-3	-5	1	0	U1	-5	∩2	1
Slope	1	0	1	1	∩2	1	∩2	-1	∩1	-4	-3	1

Species models were then applied to the mined landscape of the SHR to assess the probable occurrence of each species across the landscape. Japanese honeysuckle had the highest probable occurrence at 48 percent (73% moderate probable occurrence), with princesstree having the lowest at less than 1 percent (3% moderate probable occurrence) (Table 3-3).Overall 33 percent of the landscape was predicted to have no invasive plants, with 47 percent predicted to have one, 17 percent to have two, and 3 percent to have three or more. An example of the mapped output is given in Figure 3-1. Chinese lespedeza and Japanese honeysuckle had the highest correlation (r= -0.43), and were found on opposing areas (example Figure 3-1 B and C).

Table 3-3: Probable proportion of mined landscape invaded for six dominant invasive plant species.

Probability of Occurrence	Autumn Olive	Chinese Lespedeza	Privet	Japanese Honeysuckle	Princesstree	Sawtooth Oak
Low	0.62	0.62	0.70	0.27	0.97	0.74
Moderate	0.25	0.21	0.23	0.25	0.03	0.21
High	0.14	0.17	0.07	0.48	<0.01	0.05
Correlation	0.41	0.51	0.28	0.52	0.30	0.17

Table 3-4: Correlation between species for six dominant invasive plant species models.

	Chinese Lespedeza	Privet	Japanese Honeysuckle	Princesstree	Sawtooth Oak
Autumn Olive	-0.05	0.05	0.16	0.18	0.39
Chinese Lespedeza		-0.15	-0.42	0.13	-0.15
Privet			0.13	-0.03	0.01
Japanese Honeysuckle				-0.04	0.05
Princesstree					0.05

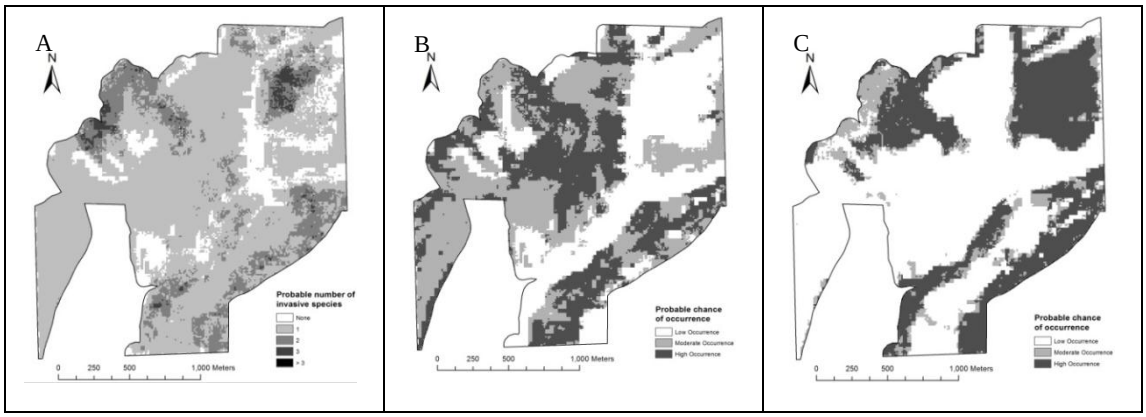

Figure 3-1: Example of mapping invasive plant distribution on reclaimed mines in the Shale Hills region, Al (A – Combined geospatial species models to give probable number of invasive species, B – Probable distribution of Chinese lespedeza, C – Probable distribution of Japanese honeysuckle).

3.5 Discussion

Geospatial modelling of these invasive plants, at this scale, is useful and offers potential for management, both in terms of identifying habitat types most at risk and areas that need management attention. Of all the geospatial variables used, distance to forest, distance to roads, NDVI in 1987 and NDVI in 2011, supplied the highest contribution to the models. This suggests that landscape disturbance and habitat characteristics (amount of forest) are greatly influencing the distribution of invasive species in the area. Although we were assessing the distribution of species invasive to the forested areas, one species that was not associated with forest was Chinese lespedeza. Chinese lespedeza is a planted species and was especially prevalent in the open areas. Chinese lespedeza and Japanese honeysuckle utilized opposing habitats in the

landscape but were ubiquitous across the area. Overall, Chinese lespedeza was found in newer mines that had a greater distance from an established forest, was closer to roads, had less forest in 1990, and a lower NDVI in 1998 and 2011. This suggested that the most recently disturbed areasare dominated by Chinese lespedeza and that this species may be competitively excluded as forest re-establishes. In our study of habitat characteristics (Chapter 2) we found Chinese lespedeza was more likely to be found in open or pine areas with higher magnesium levels in the soil and little or no midstory and downed woody debris. The models of Chinese lespedeza occurrence were not strong, however, suggesting there may be other factors that need to be considered. Chinese lespedeza has been planted since 1970 as part of reclamation plans; thiscontinues today (pers com Dr. Randall Johnson, Director, ASCM), and the effect of this planting may be what we are failing to capture. For the management of this species, increased canopy cover with a diverse forest structure seems to be the best long-term approach, but the best management practice to assist in eliminating this species from the reclaimed sites would be to remove it from allowable seeding mixtures.

Japanese honeysuckle was more likely to be found close to or within forest and with a higher NDVI in 2011, suggesting it is primarily in forested environments. This agrees with other studies that have found Japanese honeysuckle to have high shade tolerance and lower competitive abilities in open/high light environments (Miller et al. 2010). Japanese honeysuckle has been widely planted for deer and cattle forage (Dickson et al. 1978; Patterson 1976) and is now considered naturalized in upland and lowland forests as well as in forest-edge habitats (Patterson 1976; Yates et al. 2004). It is not as detrimental as some of the other alien species, but it has been shown to impact even-aged pine regeneration when established at very high densities. Given Japanese honeysuckle prevalence throughout the southeastern states, there may be scant management efficacy for its removal from the SHR.

Privet was more likely to be found close to, or within forest, close to water and on older reclaimed mines. Previous work (Chapter 2) found that privet was associated with areas with high canopy cover. Privet is considered the second most abundant invasive plant in the south and is most prevalent in the understory of bottomland hardwood forests (Merriam & Feil 2002). It can form dense stands to the exclusion of most native plants and replacement regeneration, impacting the abundance of specialist birds and diversity of native plants and bees (Wilcox & Beck 2007). We would suggest this is of management concern in the SHR and that as forest

regenerates that it would be advisable to manage for privet, particularly in the depressions and low lying areas that are more hydric and closer to water.

Autumn olive was more likely to occur within or closer to forest, closer to rivers but further from roads and with low NDVI in 1987, suggesting it is more prevalent in areas that have had fewer disturbances in recent years. Once established, it can develop intense shade which suppresses native species and can cause serious problems for native species that flourish on nitrogen poor soils (Sather & Eckardt 1987). It has been planted as part of reclamation in this area since the 1970s (pers com Dr. Randall Johnson, Director, ASMC). As this species is not at high densities throughout the larger region (Miller et al. 2012), it is very likely that removal of current infestation and no future planting activity would be very worthwhile and could possibly lead to eradication in the area.

Princesstree occurrence was more likely to be found with low NDVI in 1987 and high NDVI in 2009, with historical disturbance but no recent disturbance. Princesstree occurrence was approximately even throughout the broader region and although some active management would be useful as this species is not of major concern.

Sawtooth oak was more likely to be found within or closer to forests and further from roads. It has been planted as part of reclamation in this area since the 1970s (pers com Dr. Randall Johnson, Director, ASMC) as a wildlife species. Although sawtooth oak possesses many favourable traits, some studies have shown that it is not as hardy as some of the native oaks and may not be as long lived (Huntley & Hopkins 1979), with maturation no earlier than some of the native oaks. During the course of the field work we did not notice volunteer species, however any alien species is a risk and we would suggest that it be removed from the list of plants appropriate for reclamation.

Of interest is how the invasive plant distribution in this landscape compares to the rest of the region. Overall, all invasive species apart from Japanese honeysuckle (5%) and privet (2%) occur on an average of less than 1% of sites across the southeastern region (Miller et al. 2012). All species we modelled, apart from princesstree and sawtooth oak, showed much higher occurrence in the SHR, suggesting that there is something of concern in this area. When assessed as probable proportion of the mined landscape this discrepancy increased even further.

3.6 Literature Cited

Alston, K. & Richardson, D. (2006) The roles of habitat features, disturbance, and distance from putative source populations in structuring alien plant invasions at the urban/wildland interface on the Cape Peninsula, South Africa. Biological Conservation, 132, 183-198.

Bartuszevige, A., Gorchov, D. & Raab, L. (2006) The relative importance of landscape and community features in the invasion of an exotic shrub in a fragmented landscape. Ecography, 29, 213-222.

Bradshaw, A.D. (1984) Technology lecture: Land restoration: Now and in the future. Proceedings of the Royal Society of London. Series B, Biological Sciences (1934-1990), 223, 1-23.

Burger, J.A., Scott, D. & Mitchem, D. (2002) Field assessment of mine soil quality for establishing hardwoods in the Appalachians. Reclamation with a Purpose. (eds R. Barnhisel & M. Collins), pp. 226-240. 19th Annual Meeting, American Society of Mining and Reclamation, Lexington, KY.

Davis, M. A., Grime, J.P. & Thompson, K. (2000) Fluctuating resources in plant communities: a general theory of invasibility. Journal of Ecology, 88, 528-534.

Dickson, J.G., Segelquist, C.A. & Rogers, M.J. (1978) Establishment of Japanese honeysuckle in the Ozark Mountains. Proceedings of the Southeastern Association of Fish and Wildlife Agencies. pp. 242-245.

Donovan, M.L., Mayhew, S.L., Warren, B.E. & Stephens, V.E. (2007) An evaluation of chemical treatment and burning on the control of autumn olive (*Elaeagnus umbellate*). Report No. 3471. Michigan Department of Natural Resources, Wildlife Division. Lansing, MI

ESRI (2010) ArcGIS, Redlands, CA, USA, Environmental Systems Research Institute.

Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N. & Wickham, J. (2011) Completion of the 2006 National Land Cover Database for the Conterminous United States. Photogrammetric Engineering Remote Sensing, 77, 858-864.

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M. & Tyler, D. (2002) The national elevation dataset. Photogrammetric Engineering and Remote Sensing, 68, 5–11.

Greenburg, C.H. & Walter, S.T. (2010) Fleshy fruit removal and nutritional composition of winter-fruiting plants: a comparison of non-native invasive and native species. Natural Areas Journal, 30, 312-321.

Guisan, A., Weiss, S.B. & Weiss, A.D. (1999) GLM versus CCA spatial modeling of plant species distribution. Plant Ecology, 143, 107-122.

Gutirrez, D., Fernndez, P., Seymour, A.S. & Jordano, D. (2005) Habitat distribution models: are mutualist distributions good predictors of their associates? Ecological Applications, 15, 3-18.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. & Tatham, R.L. (2006) Multivariate Data Analysis Sixth Edition Pearson Education. New Jersey.

Hardt, R.A. & Forman, I.I. (1989) Boundary form effects on woody colonization of reclaimed surface mines. Ecology, 70, 1252-1260.

Holl, K.D. (2002) Long-term vegetation recovery on reclaimed coal surface mines in the eastern USA. Applied Ecology, 39, 960-970.

Holl, K.D. & Cairns, J.J. (2002) Monitoring ecological restoration. Handbook of Ecological Restoration. pp. 411-432. Cambridge University Press, Cambridge, UK.

Homer, C., Huang, C., Yang, L., Wylie, B. & Coan, M. (2004) Development of a 2001 National Land-Cover Database for the United States. Photogrammetric Engineering Remote Sensing, 70, 829-840.

Hosmer, D.W. & Lemeshow, S. (2000) Applied Logistic Regression. Wiley-Interscience, New York.

Huntley, C.R. & Hopkins, J.C. (1979) Establishment of Sawtooth Oak as a Mast Source for Wildlife. Wildlife Society Bulletin, 7, 253-258.

Johnson, J.E., Mitchem, D.O. & Kreh, R.E. (2003) Establishing royal paulownia on the Virginia Piedmont. New Forests, 25, 11–23.

Langdon, K.R. & Johnson, K.D. (1994) Additional notes on invasiveness of *Paulownia tomentosa* in natural areas. Natural Areas Journal, 14, 139-140.

Lemke, D., Hulme, P.E., Brown, J.A. & Tadesse, W. (2011) Distribution modelling of Japanese honeysuckle (*Lonicera japonica*) invasion in the Cumberland Plateau and Mountain Region, USA. Forest Ecology and Management, 262, 139-149.

Lockwood, J.L., Hoopes, M.F. & Marchetti, M.P. (2007) Invasion Ecology. Blackwell Publishing, Malden, Massachusetts.

Manel, S., Williams, H.C. & Ormerod, S.J. (2002) Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology, 38, 921-931.

McElfish, J.M. & Beier, A.E. (1990) Environmental regulation of coal mining: SMCRA's second decade. Environmental Law Institute, Washington, DC, USA.

Merriam, R.W. & Feil, E. (2002) The potential impact of an introduced shrub on native plant diversity and forest regeneration. Biological Invasions, 4, 369-373.

Miller, J.H., Chambliss, E.B. & Loewenstein, N.J. (2010) A field guide for the identification of invasive plants in southern forests. Gen. Tech. Rep. SRS–119. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 126 p.

Miller, J.H., Lemke, D. & Coulston, J. (2012) The Southern Forest Futures Project, Chapter 15. The Invasion of Southern Forests by Nonnative Plants: Current and Future Occupation with Impacts, Management Strategies, and Mitigation Approaches. In Press.

Oommen, T., Baise, L.G. & Vogel, R.M. (2010) Sampling bias and class imbalance in maximum-likelihood logistic regression. Mathematical Geosciences, 43, 99-120.

Patterson, D. (1976) The history and distribution of five exotic weeds in North Carolina. Southern Appalachian Botanical Society, 41, 177-180.

Phillips, S., Anderson, R. & Schapire, R. (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190, 231-259.

Ricciardi, A. (2007) Are modern biological invasions an unprecedented form of global change? Conservation Biology, 21, 329-36.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W. & Harlan, J.C. (1974) Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation. Security, 371.

Sather, N. & Eckardt, N. (1987) Element Stewardship Abstract for *Elaeagnus umbellata* (Autumn olive). Arlington, VA.

Sharp, W.C. (1977) Conservation plants for the Northeast. USDA Soil Conservation Service, Program Aid No. 1154. U.S. Government Printing Office, Washington, D.C. 40pp

Shelton, M.G. & Cain, M.D. (2002) Potential carry-over of seeds from 11 common shrub and vine competitors of loblolly and shortleaf pines. Canadian Journal of Forest Research, 32, 412-419.

Short, H.L. (1976) Composition and squirrel use of acorns of black and white oak groups. Journal of Wildlife Management, 40, 479-483.

Simberloff, D., Schmitz, D. & Brown, T. (1997) Strangers in paradise: impact and management of nonindigenous species in Florida. Island Press, Washington, DC, USA.

Stevens, D.L. & Olsen, A.R. (2004) Spatially balanced sampling of natural resources. Journal of the American Statistical Association, 99, 262-278.

Stohlgren, T.J., Chong, G.W., Schell, L.D., Rimar, K. A., Otsuki, Y., Lee, M., Kalkhan, M. A. & Villa, C. A. (2002) Assessing vulnerability to invasion by nonnative plant species at multiple spatial scales. Environmental Management, 29, 566-577.

Travis, J. & Wilterding, J. (2005) Assessment of autumn olive (*Elaeagnus umbellate*) population at Pierce Cedar Creek. Olivet College Olivet, MI.

Underwood, E.C., Ustin, S.L. & Ramirez, C.M. (2007) A comparison of spatial and spectral image resolution for mapping invasive plants in coastal California. Environmental management, 39, 63-83.

Vitousek, P.M., D'Antonio, C.M., Loope, L.L., Rejmanek, M. & Westbrooks, R. (1997) Introduced species: A significant component of human-caused global change. New Zealand Journal of Ecology, 21, 1-16.

Vogelmann, J.E., Howard, S.M., Yang, L., Larson, C.R., Wylie, B.K. & Driel, J.N. Van. (2001) Completion of the 1990's National Land Cover Data Set for the conterminous United States. Photogrammetric Engineering and Remote Sensing, 67, 650-662.

Wilcox, J. & Beck, C.W. (2007) Effects of *Ligustrum sinense* lour. (Chinese privet) on abundance and diversity of songbirds and native plants in a southeastern nature preserve. Southeastern Naturalist, 6, 535-550.

Williams, C.E. (1993) The exotic empress tree, *Paulownia tomentosa*: an invasive pest of forests? Natural Areas Journal, 13, 221-222.

Yates, E.D., Levia, D.F. & Williams, C.L. (2004) Recruitment of three non-native invasive plants into a fragmented forest in southern Illinois. Forest Ecology and Management, 190, 119-130.

Zeleznik, J.D. & Skousen, J.G. (1996) Survival of three tree species on old reclaimed surface mines in Ohio. Journal of Environmental Quality, 25, 1429-1435.

Zipper, C.E., Burger, J.A., McGrath, J., Rodrigue, J.A. & Holtzman, G.I. (2011) Forest Restoration Potential of Coal-Mined Lands in the Eastern United States. Journal of Environmental Quality, 40, 1567-1577.

CHAPTER 4.

Distribution Modelling of Japanese Honeysuckle (*Lonicera japonica*) Invasion in the Cumberland Plateau and Mountain Region, United States

Published in: *Forest Ecology and Management (2011) 262: 139-149*

Co-Authors: *P.E. Hulme, J.A. Brown, W. Tadesse*

4.1 Abstract

Predicting the potential distribution of invasive plants within a specific region is pivotal to planning effective management but is challenged by attempting to model expanding populations that are rarely at equilibrium with their environment. We adopt an ensemble modelling approach to assess the potential distribution of Japanese honeysuckle (*Lonicera japonica*), a vine invasive to forests of the Cumberland Plateau and Mountain Region in the southeast of United States. The influence of disturbance, spatial and temporal heterogeneity and other landscape characteristics were assessed by creating regional level models based on occurrence records from the United States Department of Agriculture, Forest Service, Forest Inventory and Analysis (FIA) database. Logistic regression and maximum entropy (MaxEnt) models were assessed independently and evaluated as predictive tools to test the value of presence/absence and presence only data in predicting species distributions. Ensemble models were also developed that combined the predictions of the two modelling approaches to obtain a more robust prediction. While the logistic regression and MaxEnt models were similar in their predictive ability and dominant input variables, the ensemble approach derived the best fitting model overall. The regional distribution of Japanese honeysuckle was influenced greatly by environmental conditions such as elevation, slope, and temperature with anthropogenic activity having significant, though lesser, influence. The ensemble models predict that Japanese honeysuckle has nearly reached its potential distribution. However, given the critical role of minimum temperature on Japanese honeysuckle distribution, future occupancy at higher elevations is likely to increase since

January temperatures for this region are predicted to rise by 1–4 ◦C over the next 100 years. The models also give some indication of the likely effect of land cover change on its distribution. Japanese honeysuckle tended to be associated with a high component of farming or low component of forest within the local neighbourhood. This suggests disturbed forest and/or high fragmentation has a higher invasion potential and given past trends and expected continued population growth, this disturbance and fragmentation will only increase. The models can be integrated into forest management decision support systems and assist in the development of long-term management plans, integrating the impact of potential climate and land cover change scenarios.

4.2 Introduction

The rapid anthropogenic acceleration of species introductions over the last century (Hulme et al. 2009) and subsequent consequences on economies and ecosystems (Vila et al. 2010) has led to biological invasions being recognised as a major component of global environmental change (Ricciardi 2007; Vitousek et al. 1997). The growing human population, rise in global trade, relative ease of travel and transport, and degradation of native habitats all contribute to the increase in introduction rate, establishment and spread of alien plants (D'Antonio et al. 2004; Hulme 2009; Myers and Bazely 2003). Human activities have significantly altered landscapes through the fragmentation of natural habitats and creation of agricultural and urban land. Such environmental heterogeneity is thought to play a defining role in whether ecosystems can resist alien species invasions and the rate at which an invasion process will likely occur across the landscape (Melbourne et al. 2007).

The eastern forests in the United States are experiencing significant invasions of alien plants (Martin et al. 2009). Invasive alien plants raise concerns about the continued integrity of forest ecosystems by (1) impacting native biodiversity (Reinhart et al. 2005; Standish et al. 2001) and (2) altering ecosystem processes such as nutrient cycling (Hawkes et al. 2005; Vitousek et al. 1997). Invasive plants also decrease economic returns for forest landowners by limiting natural regeneration and increasing management costs (Pimentel et al. 2005; Webster et al. 2006). Controlling invasive plants is costly and affects both the viability of the forest industry as a whole as well as ecosystem sustainability (Pimentel et al. 2005). The abundance and distribution of invasive plants are of particular concern to landowners of small non-industrial forests who often have limited resources to conduct large inventory and control projects (Byers et al. 2002).

Forests are often perceived as relatively resistant to plant invasions from all but the most shade tolerant of alien species (Martin et al. 2009; Pysek et al. 2010) and the increased infiltration of alien species into eastern forests may indicate increased vulnerability of these ecosystems due to fragmentation, disturbance and environmental heterogeneity as a result of anthropogenic activities. Understanding the drivers of plant invasions requires a landscape perspective of environmental heterogeneity in relation to the distribution of alien species. A variety of abiotic, biotic, anthropogenic, and historic variables related to environmental suitability and propagule pressure determine the distribution of invasive alien plants in a particular landscape (Kumar et al. 2006; Stohlgren et al. 1999; With & Crist 1995). Understanding the role of these variables for invasive alien plants is challenging because environmental variables are often inter-correlated, many invaders have broad ecological niches, and their distributions are often highly dynamic in both space and time (Hulme 2003).

To date, various statistical methods have been used to integrate individual species occurrence data with environmental spatial data to predict the suitability of a site for alien plant species. These have included logistic regression (Collingham et al. 2000), fuzzy envelope models (Robertson et al. 2004), genetic algorithms (Underwood et al. 2004), maximum entropy (Hoffman et al. 2008), and general additive models (Dullinger et al. 2009). These models differ in the underlying algorithms and in their requirement of species presence only data or for both presence and true absence data. A major challenge in modelling the distributions of invasive species is that it cannot be assumed that the study organism has fully spread throughout all potentially suitable sites (Peterson 2003). Thus sites from which the invasive species is known to be absent may not necessarily be unsuitable, but rather they may be yet to be colonised. Under such circumstances and especially when species prevalence is low, models using presence/absence data will tend to weigh species absence more heavily than they should while presence-only models will tend to over predict current occurrence (Vaclavik & Meentemeyer 2009). While each of these models captures only part of the distribution of a species, their errors are complimentary and the potential combination of both modelling approaches may provide the best insights into species distributions. We propose such an ensemble approach as a means to improve invasive species distribution models and to assist in developing tools for managing invasive plant species (Stohlgren et al. 2010).

4.3 Methods

4.3.1 Study Area

The Cumberland Plateau and Mountain Region (CPMR) of United States extends from northern Alabama, through Tennessee, Kentucky and into Virginia (Smalley 1979, 1982, 1984, 1986) (Figure 4-1). The CPMR covers a total area of 59,000km^2, and has one of the most diverse woody plant communities in eastern North America (Ricketts et al. 1999). Forest resources and management are a major part of the CPMR economy, particularly in rural communities. Approximately 70 percent of the land in this area is forested, with over 75 percent of this constituting hardwoods (Homer et al. 2004; USDA/FS, 2007). The CPMR is deeply dissected, with 120 – 300 m of topographic relief and frequent sandstone outcrops and bluffs (USDA 2006). Elevations range from 200 – 1200 m, with annual rainfall varying from 940 – 1900 mm. Like many of the forests in eastern North America, the native deciduous hardwood forests of the CPMR are characterised by a long history of land-use change driven by agricultural conversion and timber extraction. More recently, urban sprawl and large-scale conversion of land to intensively managed pine plantations have become major contributors to land cover change (Wear and Greis 2002). McGrath et al. (2004) found that 14 percent of native forest cover was lost between 1981 and 2000, predominantly as a result of native forest conversion to pine plantations.

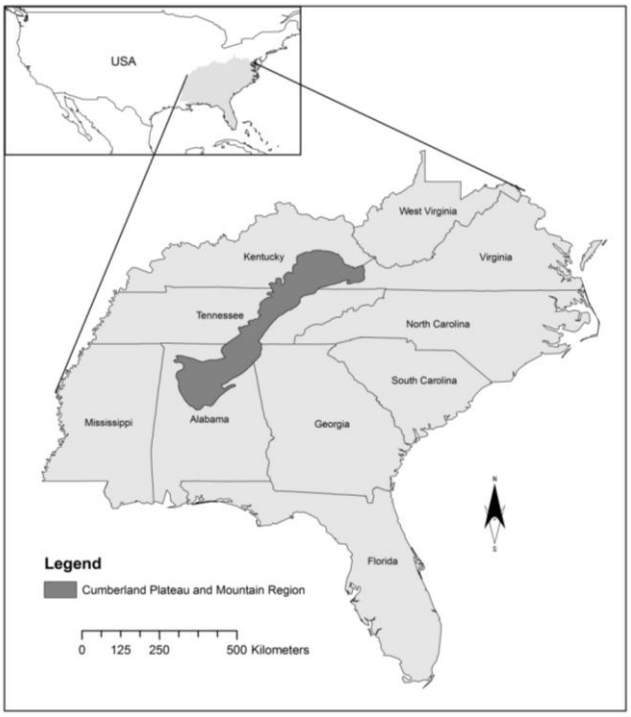

Figure 4-1: Study area location map, Cumberland Plateau and Mountain region, southeastern United States.

4.3.2 Species of Interest

Japanese honeysuckle (*Lonicera japonica* Thunb.) is the most prevalent alien species in the CPMR (USDA, 2005). It is native to Asia (Ohwi 1965) but is becoming a serious problem in many areas throughout the world including North America (Schweitzer & Larson 1999), New Zealand (Timmins & Williams 1991), southern Australia (Williams et al. 2001) and southern

Chile (Swenson et al. 1997). In North America, Japanese honeysuckle was introduced in 1806 to New York (Leatherman 1955), with the first noted escape from cultivation along the Potomac River in 1882 (U.S. National Herbarium). It was later widely planted for deer and cattle forage (Dickson et al. 1978; Patterson 1976) and is now considered naturalized in upland and lowland forests as well as in forest-edge habitats (Patterson 1976; Yates et al. 2004). It has been documented in at least 42 states with the United States and is listed as an invasive or noxious weed in several eastern states (USDA 2009). The species occurs in both open and shaded areas, with annual precipitation in invaded areas averaging 1000–1200 mm and minimum temperatures as low as −15 to −8 ◦C (Sasek & Strain 1990). Based on the current distribution in North America, its ecology, physiology, and phenotypic plasticity, the species is expected to continue to spread in eastern North America (Schierenbeck 2004). Although it is considered a widespread, naturalized weed, Japanese honeysuckle is widely available commercially throughout North America and is still planted as a cultivar in areas without a heavy frost. As recently as 1994 it had been recommended by wildlife managers for use as deer forage and cover (Dyess et al. 1994).

Japanese honeysuckle has been identified as a successful competitor and a contributor to reduced species diversity in many plant communities (Bell et al. 1988; Davison & Forman 1982). It can alter the understory and herbaceous layers of the communities it invades (Barden & Matthews 1980; Davison & Forman 1982) and ensure its continued dominance through the suppression of seedlings (Regehr 1988; Schierenbeck 2004). It also poses a threat to commercial forestry, as it has been shown to suppress seedlings of commercially valuable pine species (Cain 1991).

Japanese honeysuckle has several competitive advantages, such as a higher specific leaf area compared to native species, that allows it to succeed in a variety of light environments and rapidly exploit natural and anthropogenic disturbance of forest canopies (Baars & Kelly 1996; Schweitzer & Larson 1999). It has higher transpiration rates consistent with the limited ability of Japanese honeysuckle to tolerate drought conditions (Leatherman 1955; Schierenbeck 2004) and it has great potential for rapid dissemination of its seed through avian and deer dispersal (Bartuszevige & Gorchov 2006; Vellend 2002). Some limitations to its invasion include intensive browse by deer (Munger 2002) and limitation of pollination (Larson et al. 2002). Thus the invasion of Japanese honeysuckle will always be dynamic and influenced by multiple

external factors. Within the southeast it is still considered to be spreading (Merrian 2003) and although considered naturalized in the study area it probably has yet to reach its full potential.

4.3.3 Japanese Honeysuckle Occurrence

The Forest Service Forest Inventory and Analysis (FIA) program of the United States Department of Agriculture collects, analyses, and reports information on the status, trends and conditions of forests within the United States. Periodic surveys of the nation's forested land have occurred since 1928 (Birdsey & Schreuder 1992). Recent inventories have typically been conducted every five to seven years in the southeastern states, with approximately 20 percent of the sites assessed every year (USDA/FS 2007). In the CPMR there are 2,814 FIA sites.

An extension of the FIA database focuses on invasive plants and assesses the occurrence and density of invasive species of concern. Invasive plants are only assessed at FIA sites that are forested (1,907 sites) and only within portions of the sites that are forested. Each FIA site is made up of four circular subplots with radius of 7.32 m, thus each site encompasses an area of 673 m^2. For this study the presence of one or more plants in the site was sufficient to be scored as invaded. Although a full data collection cycle takes 5 – 7 years, we did not distinguish the timing of the samples in the analysis, since we were interested in site suitability for occupancy and the differences in sampling periods were small. The invasive plant data collection has been completed for one inventory cycle (2001–2006) with Japanese honeysuckle as the most prevalent species. Each of the 1,907 FIA sites were scored for presence or absence of Japanese honeysuckle and the geographic locations of these sites were imported into ArcGIS (ESRI 2009) and combined with landscape attributes. To limit the models to only forested area the national land cover data set was used to define the forested area of the CPMR. Japanese honeysuckle occurred in 30 percent of the FIA sites but only at 27 percent of the sites classified as forested by the land cover data (18% of FIA sites were not in areas defined as forest by the land cover data, this is due to different definitions of forest in the data sets). Although data on categorical abundance are available our aim was to compare two widely used distribution modelling techniques that use presence only or presence absence data.

4.3.4 Landscape Variables

Landscape variables were categorised into six groups: Landsat, anthropogenic, environmental, climate, land cover, and water. All variables were extracted from available digital information

including Landsat imagery, classified land cover data, roads, river, human population census and climate information, using ArcGIS and ERDAS (ERDAS Inc 2008). All variables were converted into 30 m × 30 m cells across the CPMR (Table A9-6).

Landsat imagery (TM and ETM+) was used to identify two relevant indices, Normalized Difference Vegetation Index (NDVI) (Tucker 1979) and Disturbance Index (DI) (Healey et al. 2005) at different time steps, to assess forests and forest change specifically related to disturbance. Landsat data for three time periods, 1975, 1990 and 2000, over the last 25 years were used. NDVI is numerical indicator, in this case derived from Landsat, of green vegetation, the higher the index the greater the amount of green vegetation. The DI is derived from the tassel cap index (Huang et al. 2002, Kauth & Thomas 1976) and was designed to highlight the unvegetated spectral signatures associated with stand-replacing disturbance and separate them from all other forest signatures. Due to the limited availability of scenes and constraints arising from cloud cover each data-time set was constructed with data from the growing season and within two years before and after the nominal year. Vegetation was the main characteristic of interest, specifically vegetation disturbance, thus images were selected from midsummer to allow better separation of forested and non-forested areas. Summer is also a period of high phenological stability and good spectral separation (Coppin et al. 2004). Ten Landsat scenes were required to cover the CPMR. These images were checked for accuracy of registration and reregistered if necessary (Lunetta & Elvidge 1998) before indices were calculated. These index variables (six in total) along with change in indices i.e. NDVI in 2000 – NDVI in 1990 (six variables in total) were used to examine forest cover and disturbance.

Anthropogenic factors were represented to some extent in the other landscape variables, particularly those that were a function of current land cover. In addition, seven variables were used to examine specific anthropogenic effects. Roads influence both alien plant presence (Jonathan & Jayne 2003) and forest community structure (Avon et al. 2010). We considered three road related variables: (1) distance to nearest road in 2000, (2) distance to nearest main road in 2000, and (3) density of roads in 2000. Humans are one of the main propagators of alien species and as such, the density of human settlement may greatly influence the distribution of invasive plants. We consider this with four variables, (1) population of the census block group in 2000, (2) residential presence in a 100 m buffer around the FIA site in the year 2001, (3) residential presence in a 500 m buffer in the year 2001 and (4) areas that had any residential area

in 2001 or 1992 in a 500 m buffer. These were derived from the 2000 Topologically Integrated Geographic Encoding and Referencing system (TIGER) road and census data (USBOC 2000).

We also used elevation, slope, aspect, solar radiation, and curvature, as these may all play a significant role in species distributions (Gutirrez et al. 2005). Eight variables were derived from a 30 m digital elevation model, part of the United States Geological Survey (USGS) National Elevation Dataset (Gesch et al. 2002). The digital elevation model was used to generate the following variables using ArcGIS (ESRI, 2009) Spatial Analyst Tools: (1) elevation (DEM), (2) slope in degrees (SLOPE), (3 and 4) aspect was calculated in degrees and then transformed into a linear north–south (NORTH) and east–west (EAST) gradient by performing cosine and sine transformations, respectively (Guisan et al. 1999), (5 and 6) slope was integrated using a second transformation (Piedallu & Gegout 2008) using the cosine of aspect multiplied by the sine of slope (NORTHNESS) and the sine of aspect multiplied by the sine of slope (EASTNESS), (7) hillshade (HILL), representation of solar radiation. This was calculated using an azimuth of 144◦ and altitude of 75◦, the location of the sun at midday in mid-summer and (8) curvature (CURV), is a measure of shape of the landscape, whether it is flat, convex, or concave shape. In ArcGIS the curvature tool assesses surrounding cells to calculate a curvature, with increasing positive scores representing increasing concavity. No temporal measurements were calculated for these environmental variables.

Four climate variables based on 30-year average (1971–2000) temperature and rainfall maps were obtained from a Parameter-elevation Regressions on Independent Slopes Model (PRISM) in grid format (PRISM Group, 2008). Variables used were: (1) annual rainfall (RAIN), (2) maximum average summer (June) temperature (MAXT), (3) minimum average winter (January) temperatures (MINT) and (4) mean annual temperature (AVET).

Land cover variables were extracted from the USGS National Land Cover Data (NLCD) for 1990 and 2000. They were reclassified to eight land cover types (Anderson et al. 1976): water and wetlands, residential, bare land, deciduous forest, evergreen forest, mixed forest, other natural vegetation and farming. Since each FIA site is, by definition forested, we estimated the nature of land cover surrounding each site within a 100 m and 500 m buffer. The change in forest cover between 1992 and 2001 (combined deciduous, evergreen and mixed forest), absolute proportions of forested area and farmed area were also calculated at a 100 and 500 m buffer using a moving window average.

Streams may affect the distribution and establishment of plants by influencing seed dispersal and moisture availability. Riparian areas have been shown to contain more alien plants than nearby upland areas (Stohlgren et al. 2002). Four variables (distance to stream, density of streams, occurrence of a wetland or stream within a 100 m and 500 m buffer) were derived from stream and wetland data.

Table 4-1: Description of variables (descriptive statistics for variables used for model input; * variables used in modelling)

	Variable	Variable Code	Description	Resolution	Source
Landsat	Disturbance Index for 1975*	DI75	Healey et al. 2005	900 m^2	Landsat
	Disturbance Index for 1990*	DI90	Healey et al. 2005	900 m^2	Landsat
	Disturbance Index for 2000*	DI00	Healey et al. 2005	900 m^2	Landsat
	Change in Disturbance Index between 1975 and 1990	DI90-75	Healey et al. 2005	900 m^2	Landsat
	Change in Disturbance Index between 1990 and 2000*	DI00-90	Healey et al. 2005	900 m^2	Landsat
	NDVI in 1975	NDVI75	Tucker, 1979	900 m^2	Landsat
	NDVI 1990*	NDVI90	Tucker, 1979	900 m^2	Landsat
	NDVI 2000*	NDVI00	Tucker, 1979	900 m^2	Landsat
	Difference in NDVI between 1975 and 1990	NDVI90-75	Tucker, 1979	900 m^2	Landsat
	Difference in NDVI between 1990 and 2000	NDVI00-90	Tucker, 1979	900 m^2	Landsat
Anthropogenic	Number of people per square km in 2000*	CENSUS	USBOC, 2000	Census block	Census 2000 *TIGER*
	Distance to road*	RD_DIST	USBOC, 2000	900 m^2	Census 2000 *TIGER*
	Density of roads within a km^2 area in 2000*	RD_DEN	USBOC, 2000	900 m^2	Census 2000 *TIGER*
	Distance to major road*	MRD_DIST	USBOC, 2000	900 m^2	Census 2000 *TIGER*
	Residential in 2000 or 1990 within a 500 m buffer	RES ALL	Anderson et al. 1976	900 m^2	USGS LULC
	Residential presence within a 100 m buffer in 2000*	RES100	Anderson et al. 1976	900 m^2	USGS LULC
	Residential presence within a 500 m buffer in 2000*	RES500	Anderson et al. 1976	900 m^2	USGS LULC
Climate	Minimum temperature from a 30year average*	MINT	PRISM Group 2008	900 m^2	PRISM
	Maximum temperature from a 30year average	MAXT	PRISM Group 2008	900 m^2	PRISM
	Average temperature from a 30year average	AVET	PRISM Group 2008	900 m^2	PRISM
	Average yearly rainfall from a 30year average*	RAIN	PRISM Group 2008	900 m^2	PRISM

Table 4-1 continued: Description of variables (descriptive statistics for variables used for model input)

	Variable	Variable Code	Description	Resolution	Source
Environmental	North	NORTH	Guisan et al. 1999	900 m^2	USGS National Elevation Dataset
	East	EAST	Guisan et al. 1999	900 m^2	USGS National Elevation Dataset
	Northness*	NORTHNESS	Piedallu and Gégout, 2008	900 m^2	USGS National Elevation Dataset
	Eastness*	EASTNESS	Piedallu and Gégout, 2008	900 m^2	USGS National Elevation Dataset
	Slope*	SLOPE	ESRI, 2009	900 m^2	USGS National Elevation Dataset
	Hillshade*	HILL	ESRI, 2009	900 m^2	USGS National Elevation Dataset
	Curvature	CURV	ESRI, 2009	900 m^2	USGS National Elevation Dataset
	Elevation*	DEM	Gesch et al. 2002	900 m^2	USGS National Elevation Dataset
Land Cover	Change in forest between 2000 and 1990 within a 100 m buffer	FC100	Anderson et al. 1976	900 m^2	USGS LULC
	Change in forest between 2000 and 1990 within a 500 m buffer*	FC500	Anderson et al. 1976	900 m^2	USGS LULC
	Proportion of forest in 2000 within a 100 m buffer*	F00 100	Anderson et al. 1976	900 m^2	USGS LULC
	Proportion of forest in 2000 within a 500 m buffer	F00 500	Anderson et al. 1976	900 m^2	USGS LULC
	Proportion of farming in 2000 within a 100 m buffer	FARM100	Anderson et al. 1976	900 m^2	USGS LULC
	Proportion of farming in 2000 within a 500 m buffer*	FARM500	Anderson et al. 1976	900 m^2	USGS LULC
	Categorical land use in 1990 based on Andersons groupings*	LULC90	Anderson et al. 1976	900 m^2	USGS LULC
	Categorical land use in 2000 based on Andersons groupings*	LULC00	Anderson et al. 1976	900 m^2	USGS LULC
Water	Distance of from a stream*	RIV DIS	seamless.usgs.gov	900 m^2	USGS
	Density of streams within a km2 area*	RIV_DEN	seamless.usgs.gov	900 m^2	USGS
	Occurrence of a wetland or stream within a 100 m*	WATER100	Anderson et al. 1976	900 m^2	USGS LULC
	Occurrence of a wetland or stream within a 500 m*	WATER500	Anderson et al. 1976	900 m^2	USGS LULC

This initial selection of variables totalled 41 (Table 4-1). The set was reduced using exploratory data analysis to remove variables that were highly correlated with another variable. Pearson's correlations were used to assess the independence of these variables. For variables with high correlation ($r > 0.71$) only one variable was selected for input into further models, the highest remaining correlation was 0.69 ($r^2 = 0.48$) (Tables A9-7 to A9-12). All input variables were mapped; NDVI75 and NDVI90-75 showed strong blocks based on scenes, an artefact of

instrumentation, and thus were not suitable for use in further analysis. This left a set of 28 variables (Table 4-1; *).

Descriptive statistics for the 28 variables were calculated for both the FIA and CPMR data sets to determine if FIA data can be extrapolated to the entire CPMR. The FIA points had a mean within one standard deviation of the mean for the forested area of the CPMR for all variables. All but two variables had means within 0.2 standard deviations. The mean for the CPMR for farming within 500 m was within 0.56 standard deviations of the FIA data and forest within 100 m was within 0.37 standard deviations. In both cases the maximum and minimum were very similar suggesting that although there was some variation in the means they still represent the full range of the CPMR. Overall this shows that FIA data are a good representation of the CPMR.

4.3.5 Models

Two modelling techniques, logistic regression (Hosmer & Lemeshow 2000) and maximum entropy (Phillips et al. 2006), were used to develop probability envelopes for Japanese honeysuckle occurrence. The important difference between the two techniques is that logistic regression uses information on both presence and absence to estimate a predictive linear model, whereas maximum entropy (MaxEnt) uses information from presence only and is a nonparametric approach. A drawback of maximum entropy is that it may give very large predicted values for environmental conditions outside the range present in the study area (Phillips et al. 2006). This is due to the exponential modelling process which can lead to the values outside of the modelled range which may be very poorly represented as there is no upper limit to the predicted values. This is not an issue in this study since we are only predicting distribution from within the range that the data has been collected. The distribution of Japanese honeysuckle was modelled using each group of variables (Landsat, anthropogenic, environmental, land cover, water and climate) separately. These "sub-models" were built using each of the two techniques, logistic regression and maximum entropy. Using only variables selected in the final sub-model for each variable group a final best model was determined. Logistic regression modelling was conducted using SAS (SAS Institute 2009) and maximum entropy modelling was performed using a specialised package of MaxEnt (Phillips et al. 2006). Logistic regression models were derived for each data group using a stepwise regression method with Akaike's Information Criterion (AIC) (Akaike 1974) as the selection criterion. Maximum

entropy models were derived for each data group and variables with little or no impact were removed using a manual backward selection method. Impact on the model was measured as percent contribution, jack-knife test on gain, and influence on area under the receiver-operating curve (ROC).

The omission rate, Cohen's Kappa and Area under the ROC (AUC) were used to assess the reliability and validity of models. The omission rate is the false negative or the proportion of sites where the species was present but the model predicted absence. The Kappa statistic is the chance-corrected proportional agreement, where possible values range from +1 (perfect agreement) to −1 (complete disagreement), via 0 (no agreement above that expected by chance) and where 0.01–0.2 = slight agreement, 0.21–0.4 = fair agreement, 0.41–0.6 = moderate agreement, 0.61–0.8 = substantial agreement, and 0.81–1 = almost perfect agreement (Landis & Koch 1977). To calculate these accuracy measures, a cut-off criterion is required to convert continuous model predictions to binary classifications. We used a threshold value that maximized the sum of sensitivity and specificity. For each potential cut-off at 0.001 intervals, for each model, the sensitivity (the proportion of actual positives which are correctly identified) and the specificity (the proportion of negatives which are correctly identified) were calculated. Sensitivity and specificity for each potential cut-off were added together and the cut-off with the greatest combined number was selected for further work. This has the advantage of giving equal weights to the probability of success of both presences and absences (Manel et al. 2001). This is one of the most appropriate methods to correctly derive a binary variable from continuous probabilities when species presence–absence distribution data are unbalanced (Jimenez-Valverde & Lobo 2006; Liu et al. 2005). AUC provides a single measure of model performance, independent of any particular choice of threshold but is sensitive to the method in which absences in the evaluation data are selected (Lobo et al. 2008). We used the following classes of AUC to assess model performance: 0.50–0.75 = fair, 0.75–0.92 = good, 0.92–0.97 = very good, and 0.97–1.00 = excellent (Hosmer & Lemeshow 2000).

Maps were created in ArcGIS. Rasters were imported from MaxEnt and the raster calculator was used to apply the logistic regression model. Initial maps were continuous rasters that were reclassified into binary rasters based on the cut-off values determined by the maximized sum of sensitivity and specificity. A final map was developed by combining the best binary maps from each model to assess the differences in spatial representation for both models. As FIA data are only for forested lands the final models were masked by the forest cover from the 2001 land

cover data to show only forested lands. FIA defines forested sites as a site with more than 10 percent stocking, this is a lower rate than the national land cover data. However, the national land cover data are the only available regional coverage of forest cover. For comparison between models and FIA data the proportion of FIA plots in forest as defined by the land cover data were calculated. Limited evaluation of climate change and land cover change was conducted on the models. Assessment of urbanization change was done by evaluating the national land cover data sets for 1992 and 2001. Based on Dale et al. (2009) evaluation that the CPMR January temperatures will rise by 1–4 ◦C, the impact of climate change was assessed by applying a more extreme 4◦ increase in the minimum temperature.

4.3.6 Data Selection

To assess models, test and training occurrence data were split spatially with 30% in test and 70% in training datasets. Data points with Japanese honeysuckle presence that were immediately adjacent to each other were placed in different groups and remaining data were randomly assigned to test and training groups to give a 30/70 ratio. Japanese honeysuckle had 405 (21%) occurrence points in the training data and 174 (9%) occurrence points in the test. There were 930 (49%) true-absence points (that is, sites that had been surveyed for Japanese honeysuckle and it was found to be absent) in the training data and 398 (21%) in the test data. In comparison MaxEnt does not specifically use "absence" data, but rather uses information from the overall landscape. When the MaxEnt models were fitted this information came from the 1,907 data points. Separate logistic regression and MaxEnt sub-models were run for each of the six different set of independent variables: environmental, climate, land-use, anthropogenic, Landsat and water. Subsequently, those variables identified as significant in these separate models were used in a single "composite" model to assess the relative importance of these key variables.

4.3.7 Ensemble Modelling

While logistic regression and MaxEnt models may be compared individually to select the best overall model for particular datasets, methods to combine the two models have the potential to reduce the uncertainty associated with any one particular algorithm. A number of approaches have been proposed for combining outputs of individual models for ensemble predictions (Araujo & New 2007). Here we adopt a consensus approach, averaging the models, that identifies the most consistent patterns of occupancy. For example if the MaxEnt model only just

classifies an area as having the occurrence of Japanese honeysuckle but Logistic regression does not, it will be excluded from the ensemble model. However, if the MaxEnt models strongly classify an area as having occurrence and logistic regression weakly rejects it, it will show occurrence in the ensemble model. We adopt two consensus approaches which involve integrating the top performing logistic regression and MaxEnt models. The first is unweighted and averages the logistic regression and MaxEnt probabilities to derive a combined probability of occupancy between 0 and 1. The threshold probability for predicted presence in the ensemble model is simply the average threshold used in each of the two different models. The second approach is weighted and calculates a new threshold probability as previously used by maximizing the sum of sensitivity and specificity. Where this threshold value departs from the arithmetic average of the two model thresholds, it will weigh the probability scores of one model more than another. The omission rate, Cohen's Kappa and AUC were used to assess the reliability and validity of the weighted and unweighted ensemble models.

4.4 Results

All sub-models from both modelling techniques provided better than random predictions of potential Japanese honeysuckle occurrence but differed considerably in their goodness of fit in the following order of decreasing model performance: environmental > climate > land-use > anthropogenic > Landsat > water (Table 4-2). Only the first three sub-models revealed at least moderate agreement (Kappa statistic) and good performance (AUC). As a result the composite models that used the best fitting variables found in the sub-models were strongly biased towards environmental variables and Landsat variables were not included (

Table 4-3). However, all variables included in the composite models had the same sign as in the sub-group models, which suggest robust relationships. The similarity between the logistic regression and MaxEnt models in terms of the variables selected as important also supports confidence in the reliability of these outputs. Furthermore, as might be expected the composite models using the most significant variables were the top performing models examined with elevation explaining more than 50% of the variance reflecting the lowland distribution of Japanese honeysuckle.

The composite logistic regression model indicated that Japanese honeysuckle presence was more likely to occur at warmer sites, close to roads, and at low elevations where the land surrounding the FIA site was reasonably flat, farmed and supported a relatively high human population. The composite MaxEnt model was similar though the influence of climate on occurrence was much stronger and the role of human population density and farming was weaker with occurrence better reflecting low forest cover. The composite MaxEnt model was more parsimonious using only five variables compared to nine for the logistic regression. But while the omission rates were lower for the MaxEnt model, both Kappa and AUC were very similar for both models (Table 4-2). The composite logistic regression model predicted a lower level of occupancy (only 12% of forested areas, Figure 4-2A) while the composite MaxEnt model predicted a higher level (37% of forested areas, Figure 4-2B) than the FIA dataset (27%). Across the CPMR, the logistic regression and MaxEnt composite models similarly predicted presence or absence in 75 percent of sites. For 25 percent of cases the model predictions did not coincide and the MaxEnt models predicted presence in 99.6 percent of these cases (Figure 4-3). The two composite models were combined using ensemble modelling. The ensemble models gave the best assessment statistics of any of the models with little difference between the weighted and unweighted ensembles due to the similarity of the underlying thresholds (Table 4-2). The increase in goodness of fit found by using an ensemble approach was relatively minor due to the similarity in the variables included in the composite logistic regression and MaxEnt models and this again suggest fairly robust predictions (Figure 4-2C). In both ensemble models levels of predicted occupancy were close to FIA occupancy. The unweighted occupation was 31 percent and the weighted occupation was 28 percent. To assess the potential impact of climate change the most extreme scenario given by Dale et al. (2009) showed the potential to increase substantially, to approximately half of the forested areas (logistic regression composite 18%, MaxEnt composite 56%, weighted ensemble 48%, unweighted ensemble 50%).

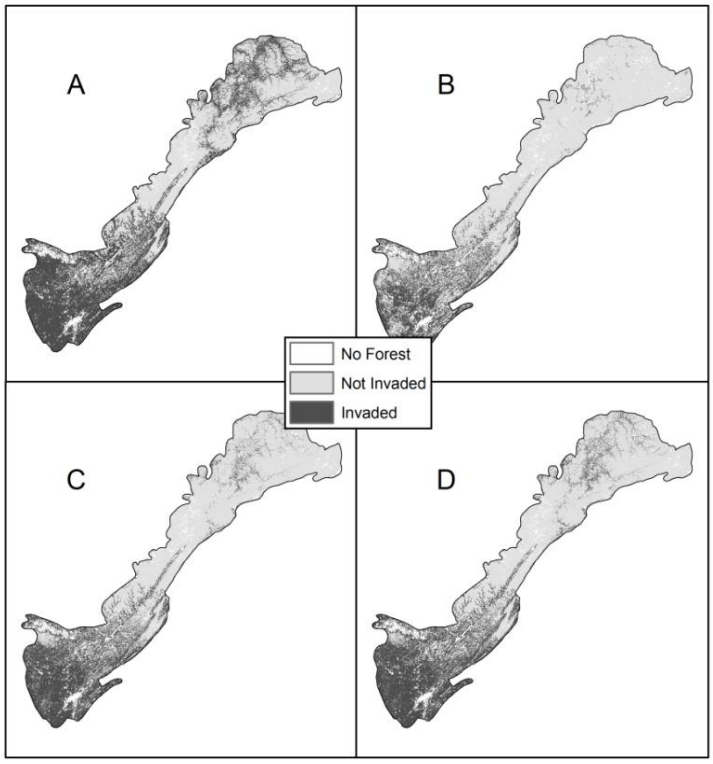

Figure 4-2: Spatial representation of model predictions for Japanese honeysuckle (A – MaxEnt, B – Logistic regression, C –Weighted ensemble, D – Unweighted ensemble).

Table 4-2: Cut-off (defined as maximum sensitivity plus specificity) and accuracy assessment for Japanese honeysuckle (bold denotes models with omission rates of less than 0.3 as an average of test and training data, models with Kappa coefficient of moderate agreement as an average of test and training data (0.41+), models with good or better AUC as an average of test and training data (0.75+)), proportion of occurrence is the proportion of forested area that has potential for invasion.

Model	Group	Threshold	Omission Rate		Kappa Coefficient		AUC		Proportion of Occurrence	Model Ranking
			train	test	train	test	train	test		
L	Landsat	0.27	0.34	0.31	0.28	0.3	0.71	0.74	0.31	14
M	Landsat	0.53	0.44	0.4	0.27	0.39	0.74	0.77	0.24	12
L	Anthro	0.28	0.29	0.26	0.28	0.29	0.77	0.76	0.54	10.5
M	Anthro	0.50	0.33	0.33	0.31	0.40	0.73	0.75	0.39	12
L	Enviro	0.25	0.19	0.16	0.49	0.48	0.84	0.85	0.44	5
M	Enviro	0.49	0.27	0.26	0.51	0.55	0.84	0.86	0.32	6.5
L	Climate	0.30	0.24	0.18	0.47	0.46	0.79	0.79	0.41	8
M	Climate	0.48	0.25	0.18	0.48	0.52	0.81	0.81	0.37	6.5
L	Land cover	0.29	0.33	0.25	0.41	0.45	0.78	0.79	0.19	9
M	Land cover	0.53	0.44	0.46	0.42	0.37	0.77	0.77	0.03	10.5
L	Water	0.31	0.50	0.41	0.25	0.3	0.65	0.66	0.59	15
M	Water	0.57	0.89	0.91	0.11	0.04	0.64	0.68	0.05	16
L	Composite	0.35	0.26	0.24	0.55	0.53	0.86	0.89	0.12	4
M	Composite	0.42	0.20	0.18	0.50	0.55	0.87	0.87	0.37	3
Unweighted Ensemble		0.38	0.24	0.21	0.55	0.54	0.89	0.88	0.31	1
Weighted Ensemble		0.43	0.27	0.24	0.57	0.54	0.89	0.88	0.28	2

Table 4-3: Contribution of each variable to the final Japanese honeysuckle models (- negative, + positive, ∩ or U for binomial relationship), dominant variables given in bold.

	Water			Land Cover				Climate		Environmental				Anthropogenic				Landsat		
	WATER 500	WATER 100	RIV DIST	LULC 00	FC 500	FARM 500	F00 100	RAIN	MINT	SLOPE	EAST NESS	DEM		MRD DIST	RD DIST	RD DEN	CEN	NDVI 90	NDVI 00	DI00
L										(-)19	(+)3	(-)78						(-)30	(-)70	
M										(-)13		(-)87						(∩)39	(-)49	(+)12
L								(-)15	(+)86					(-)63	(-)20		(+)17			
M								(∩)91	(+)9					(U)65		(∩)23	(+)12			
L				4	(+)38	(+)58														
M					(+)35	(+)38	(-)27													
L	(-)76	(-)16	(+)8																	
M	(+)96	(+)4																		
	Combination of all final variables used in the six groups above, 'composite' models																			
L	(-)4					(+)15			(+)4	(-)9	(+)5	(-)53		(-)2	(-)4		(+)4			
M					(-)10				(+)22	(-)6		(-)58		(U)4						

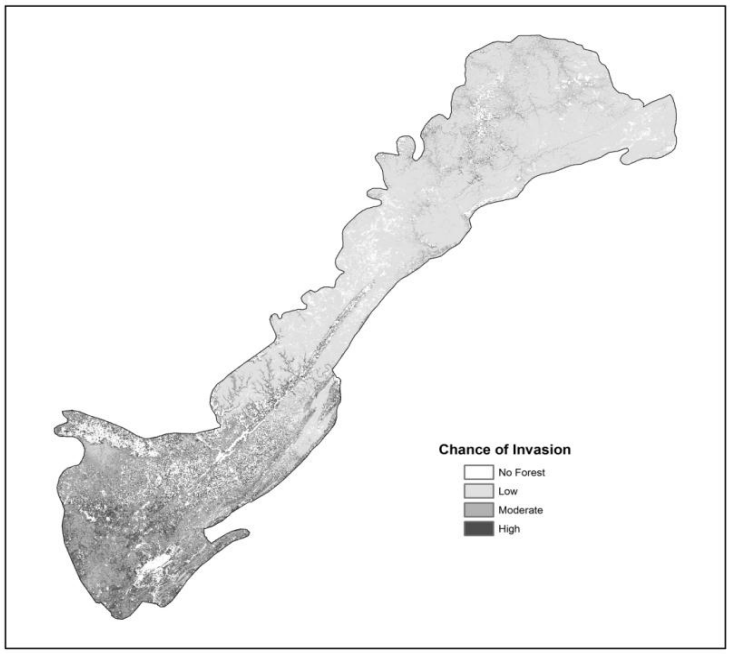

Figure 4-3: Spatial representation of binary combination of the MaxEnt and logistic regression models (low probability = absent in both models, moderate probability = presence predicted by only one model, high probability = presence predicted by both models)

4.5 Discussion

Predicting the future distribution of invasive plants is pivotal to planning effective forest management but is challenged by the fact that expanding populations are rarely at equilibrium with their environment due to progression of invasion and ongoing changes in the invaded landscape (land cover change, climate change). By combining insights from two different modelling techniques, this study has overcome some of these limitations and produced more robust estimates of the key environmental drivers underpinning the distribution of Japanese honeysuckle in the CPMR. It shows within the current forested landscape that Japanese honeysuckle is nearing its potential maximum distribution, with most of the lowland area vulnerable to invasion. Since this species is known to reduce species diversity (Bell et al. 1988; Davison & Forman 1982), alter understory conditions (Barden & Matthews 1980; Davison & Forman 1982) and to suppress seedlings (Cain 1991; Regehr 1988; Schierenbeck 2004), it is useful for managers to be able to assess the invasion probability of any given forest stand and identify areas of most issue for management action.

Our models showed that regional distribution of Japanese honeysuckle was influenced greatly by environmental conditions such as elevation, slope and temperature (86% for MaxEnt and 75%

for logistic regression) with anthropogenic activity having some influence on the distribution (14% for MaxEnt and 25% or logistic regression). These predictions assume a status quo in the environmental variables but in the future, climate and land cover change may further exacerbate this threat. All models revealed a higher chance of Japanese honeysuckle occurrence with higher minimum temperatures. Climate change is likely to alter the spatial pattern of habitats for both native and alien plants (Hellmann et al. 2008; Walther et al. 2009). Using models calibrated with current data to predict the expected future distributions of potential habitats under scenarios of environmental change is valuable for forecasting the ecological consequences of climate change. It has been predicted that the January temperatures for the CPMR will rise by 1–4 ◦C over the next 100 years (Dale et al. 2009). As minimum temperatures rise, the probability of Japanese honeysuckle colonising higher elevation will increase. Application of a 4◦ increase in minimum temperature showed the potential of Japanese honeysuckle to increase approximately 50 percent of the forest.

The models also give some indication of the likely effect of land cover change on Japanese honeysuckle distribution. In the CPMR, Japanese honeysuckle tended to be associated with neighbouring areas with a high component of farming or low component of forest. These variables are good indicators of disturbed forest and fragmentation. These areas have a high invasion potential because disturbed forests are more susceptible to the establishment of alien plants and edge areas are often the source of invasion (Cadenasso & Pickett, 2001; Yates et al. 2004). While both our models included distance to main road as an important variable, caution should be applied in extrapolating this to future scenarios. Japanese honeysuckle is strongly associated with roadsides and while this will reflect the disturbed nature of these habitats, the species has also been widely planted along roads for erosion control (Hardt, 1986). In the absence of further deliberate planting, the effect of further encroachment into the forest by roads may be less dramatic than in the past. From 1992 to 2001 residential area increased by four fold according to the national land cover database and these rates are likely to continue. With the continuing increase in population and associated development, the edge and disturbed forest habitat preferred by Japanese honeysuckle will also increase thus allowing for an increased distribution.

Although fine-scale surveys of invasive species provide insights into the factors driving species occurrence and abundance (Truscott et al. 2008; Affre et al. 2010), such approaches are not feasible for large regions such as the CPMR (59,000 km^2). Under these circumstances, model

selection and mapping occurrence predictions is the final step in species distribution modelling. In some cases it is important to give an accurate current occurrence prediction; in others it is important to assess potential distribution. Model selection can be undertaken in a number of ways, we used two; use of evaluation statistics and ensemble modelling. The composite MaxEnt and logistic regression models had very similar kappa and AUC. However, the MaxEnt models had a lower omission rate and the logistic regression predicted a much smaller area of occurrence. Both methods give valuable and complementary information on species distributions. Logistic regression only uses the parametric relationships and integrates both presence and absence data. While the logistic regression models may be best representative of area with very high probability of invasion, the MaxEnt models may be more useful for potential distribution. The use of both models together can assist in identifying areas of highest invasion potential and the best allocation of resources. Recent literature suggests that an ensemble approach may be more useful than a single model approach (Araujo & New 2007; Stohlgren et al. 2010) and this was confirmed in the present study. Nevertheless, these models remain correlative indicators of potential distribution that may underestimate demographic drivers of range expansion such as propagule pressure which may only be revealed by experiments along environmental gradients (e.g. Willis & Hulme 2002; Ross et al. 2008).

4.6 Conclusions

In this study we attempted to understand the factors responsible for shaping Japanese honeysuckle distributions in the CPMR and to identify the areas most vulnerable to invasion. To accomplish this we used different types of modelling techniques and assessed their value, both ecologically and statistically. Models such as those developed by this research can be used as tools for landscape management, forest stand assessment, and long term forest monitoring programs. In the context of Japanese honeysuckle we have been able to identify areas of probable invasion. One of the greatest benefits of large-scale GIS models is that they can outline the main characteristics of these species' distribution areas and be used to predict environmental favourability in areas where their distribution is less documented (Barbosa et al. 2009). In this paper we took just 1,907 forested survey locations and estimated the distribution for over 40,000 km^2. Models such as these can be integrated into forest management decision support systems (Ducheyne et al. 2006) and assist in the development of long term management plans, integrating the impact of potential climate change scenario (Rose & Burton 2009).

4.7 Literature Cited

Affre, L., Suehs, C.M., Charpentier, S., Vila, M., Brundu, G., Lambdon, P., Traveset, A. & Hulme, P.E. (2010) Consistency on the habitat degree of invasion for three invasive plant species across Mediterranean islands. Biological Invasions, 12, 2537–2548.

Akaike, H. (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19, 716–723.

Anderson, J.R., Hardy, E.E., Roach, J.T. & Witmer, R.E. (1976) A land use and land cover classification system for use with remote sensor data. United States Geological Survey Professional Paper 964, United States Government Printing Office, Washington, D.C.

Araujo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. Trends in Ecology and Evolution, 22, 42–47.

Avon, C., Berges, L., Dumas, Y. & Dupouey, J.L. (2010) Does the effect of forest roads extend a few metres or more into the adjacent forest? A study on understory plant diversity in managed oak stands. Forest Ecology and Management, 259, 1546–1555.

Baars, R. & Kelly, D. (1996) Survival and growth responses of native and introduced vines in New Zealand to light availability. New Zealand Journal of Botany, 34, 389–400.

Barbosa, A.M., Real, R. & Mario Vargas, J. (2009) Transferability of environmental favourability models in geographic space: the case of the Iberian desman (*Galemys pyrenaicus*) in Portugal and Spain. Ecological Modelling, 220, 747–754.

Barden, L.S. & Matthews, J.F. (1980) Change in abundance of honeysuckle (*Lonicera japonica*) and other ground flora after prescribed burning of a piedmont pine forest. Castanea, 45, 257–260.

Bartuszevige, A. & Gorchov, D. (2006) Avian seed dispersal of an invasive shrub. Biological Invasions, 8, 1013–1022.

Bell, D.J., Forseth, I.N. & Teramura, A.Z. (1988) Field water relations of three temperate vines. Oecologia, 74, 537–545.

Birdsey, R.A. & Schreuder, H.T. (1992) An overview of forest inventory and analysis estimation procedures in the eastern United States – with an emphasis on the components of change. USDA Forest Service General Technical Report RM-214.

Byers, J.E., Reichard, S., Randall, J.M., Parker, I.M., Smith, C.S., Lonsdale, W.M., Atkinson, I.A.E., Seastedt, T.R., Williamson, M., Chornesky, E. & Hayes, D. (2002) Directing research to reduce the impacts of nonindigenous species. Conservation Biology, 16, 630–640.

Cadenasso, M.L. & Pickett, S.T.A. (2001) Effect of edge structure on the flux of species into forest interiors. Conservation Biology, 15, 91–97.

Cain, M.D. (1991) The influence of woody and herbaceous competition on early growth of naturally regenerated loblolly and shortleaf pines. Southern Journal of Applied Forestry, 15, 179–185.

Collingham, Y.C., Wadsworth, R.A., Huntley, B. & Hulme, P.E. (2000) Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. Journal of Applied Ecology, 37, 13–27.

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B. & Lambin, E. (2004) Review Article digital change detection methods in ecosystem monitoring: a review. International Journal of Remote Sensing, 25, 1565–1596.

Dale, V., Lannom, K., Tharp, M., Hodges, D. & Fogel, J. (2009) Effects of climate change, land-use change, and invasive species on the ecology of the Cumberland forests. Canadian Journal of Forest Research, 39, 467–480.

D'Antonio, C.M., Jackson, N.E., Horvitz, C.C. & Hedberg, R. (2004) Invasive plants in wildland ecosystems: merging the study of invasion processes with management needs. Frontiers in Ecology and the Environment, 2, 513–521.

Davison, S.E. & Forman, R.T.T. (1982) Herb and shrub dynamics in a mature oak forest: a thirty-year study. Bulletin of the Torrey Botanical Club, 109, 64–73.

Dickson, J.G., Segelquist, C.A. & Rogers, M.J. (1978) Establishment of Japanese honeysuckle in the Ozark Mountains. In: Proceedings of the Southeastern Association of Fish and Wildlife Agencies 242–245. Southeastern Association of Fish and Wildlife Agencies , Tallahassee, FL, US.

Ducheyne, E.I., De Wulf, R.R. & De Baets, B. (2006). A spatial approach to forest-management optimization: linking GIS and multiple objective genetic algorithms. International Journal of Geographical Information Science, 20, 917–928.

Dullinger, S., Kleinbauer, I., Peterseil, J., Smolik, M. & Essl, F. (2009). Niche based distribution modelling of an invasive alien plant: effects of population status, propagule pressure and invasion history. Biological Invasions, 11, 2401–2414.

Dyess, J.G., Causey, M.K., Striblin, H.L. & Lockaby, B.G. (1994) Effects of fertilization on production and quality of Japanese honeysuckle. Southern Journal of Applied Forestry, 18, 68–71.

ERDAS Inc. (2008) Imagine 9.2, 5051 Peachtree Corners Circle, Norcross, GA.

ESRI (2009) ArcGIS, Redlands, CA, USA, Environmental Systems Research Institute.

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M. & Tyler, D. (2002) The national elevation dataset. Photogrammetric Engineering and Remote Sensing, 68, 5–11.

Guisan, A., Weiss, S.B. & Weiss, A.D. (1999) GLM versus CCA spatial modeling of plant species distribution. Plant Ecology, 143, 107–122.

Gutirrez, D., Fernndez, P., Seymour, A.S. & Jordano, D. (2005) Habitat distribution models: are mutualist distributions good predictors of their associates? Ecological Applications, 15, 3–18.

Hardt, R.A. (1986) Japanese honeysuckle: from "one of the best" to ruthless pest. Arnoldia 46, 27–34.

Hawkes, C.V., Wren, I.F., Herman, D.J. & Firestone, M.K. (2005) Plant invasion alters nitrogen cycling by modifying the soil nitrifying community. Ecology Letters, 8, 976–985.

Healey, S.P., Cohen, W.B., Zhiqiang, Y. & Krankina, O.N. (2005) Comparison of Tasseled Cap-based Landsat data structures for use in forest disturbance detection. Remote Sensing of Environment, 97, 301–310.

Hellmann, J.J., Byers, J.E., Bierwagen, B.G. & Dukes, J.S. (2008) Five potential consequences of climate change for invasive species. Conservation Biology, 22, 534–543.

Hoffman, J.D., Narumalani, S., Mishra, D.R., Merani, P. & Wilson, R.G. (2008) Predicting potential occurrence and spread of invasive plant species along the North Platte River, Nebraska. Invasive Plant Science, 1, 359–367.

Homer, C., Huang, C.Q., Yang, L.M., Wylie, B. & Coan, M. (2004) Development of a 2001 national land-cover database for the United States. Photogrammetric Engineering and Remote Sensing, 70, 829–840.

Hosmer, D.W. & Lemeshow, S. (2000) Applied Logistic Regression. Wiley-Interscience, New York.

Hulme, P.E. 2003. Biological invasions: winning the science battles but losing the conservation war? Oryx, 37, 178–193.

Huang, C., Wylie, B., Yang, L., Homer, C. & Zylstra, G. (2002) Derivation of a tasselled cap transformation based on Landsat 7 at-satellite reflectance. International Journal of Remote Sensing, 23, 1741–1748.

Hulme, P.E. (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalisation. Journal of Applied Ecology, 46, 10–18.

Hulme, P.E, Pysek, P., Nentwig, W. & Vila, M. (2009) Will threat of biological invasions unite the European Union? Science, 324, 40–41.

Jimenez-Valverde, A. & Lobo, J.M. (2006) The ghost of unbalanced species distribution data in geographical model predictions. Diversity and Distributions, 12, 521–524.

Jonathan, L.G. & Jayne, B. (2003) Roads as conduits for exotic plant invasions in a semiarid landscape. Conservation Biology, 17, 420–432.

Kauth, R. J. & G. S. Thomas. (1976) The tasseled cap –a graphic description of the spectral-temporal development of agricultural crops as seen in Landsat. Pages 41–51 in Symposium on Machine Processing of Remotely Sensed Data,. LARS, Purdue University, West Lafayette, Indiana.

Kumar, S., Stohlgren, T.J. & Chong, G.W. (2006) Spatial heterogeneity influences native and alien plant species richness. Ecology, 87, 3186–3199.

Landis, J.R. & Koch, G.G. (1977) The measurement of observer agreement for categorical data. Biometrics, 33, 157–174.

Larson, K.C., Fowler, S.P. & Walker, J.C. (2002) Lack of pollinators limits fruit set in the exotic Lonicera japonica. American Midland Naturalist, 148, 54–60.

Leatherman, A.D. (1955) Ecological life-history of *Lonicera japonica* Thunb. Dissertation. University of Tennessee, Knoxville, TN, USA.

Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography, 28, 385–393.

Lobo, J.M., Jimenez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography, 17, 145–151.

Lunetta, R.S. & Elvidge, C.D. (1998) Remote Sensing Change Detection. Ann Arbor Press, MI.

Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence/absence models in ecology: the need to account for prevalence. Journal of Applied Ecology, 38, 921–931.

Martin, M., Tipping, P. & Sickman, J. (2009) Invasion by an exotic tree alters above and belowground ecosystem components. Biological Invasions, 11, 1883–1894.

McGrath, D.A., Evans, J.P., Smith, C.K., Haskell, D.G., Pelkey, N.W., Gottfried, R.R. & Brockett, C.D. (2004) Mapping land-use change and monitoring the impacts of hardwood to pine conversion on the Southern Cumberland Plateau in Tennessee. Earth Interactions, 8, 1–24.

Melbourne, B.A., Cornell, H.V., Davies, K.F., Dugaw, C.J., Elmendorf, S., Freestone, A.L., Hall, R.J., Harrison, S., Hastings, A., Holland, M., Holyoak, M., Lambrinos, J., Moore, K. & Yokomizo, H. (2007) Invasion in a heterogeneous world: resistance, coexistence or hostile takeover? Ecology Letters, 10, 77–94.

Merriam, R.W. (2003) The abundance, distribution and edge associations of six nonindigenous, harmful plants across North Carolina. Journal of the Torrey Botanical Society, 130, 283–291.

Munger, G.T. (2002) *Lonicera japonica*. In: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory. Fire Effects Information System. http://www.fs.fed.us/database/feis/plants/vine/lonjap/all.html, 43 p.

Myers, J.H. & Bazely, D.R. (2003) Ecology and Control of Introduced Plants. Cambridge University.

Ohwi, J. (1965) Flora of Japan. Smithsonian Institution, Washington, D.C.

Patterson, D. 1976. The history and distribution of five exotic weeds in North Carolina. Southern Appalachian Botanical Society, 41, 177–180.

Peterson, A.T. (2003) Predicting the geography of species' invasions via ecological niche modeling. Quarterly Review of Biology, 78, 419–433.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190, 231–259.

Piedallu, C. & Gegout, J. (2008) Efficient assessment of topographic solar radiation to improve plant distribution models. Agricultural and Forest Meteorology, 148, 1696–1706.

Pimentel, D., Zuniga, R. & Morrison, D. (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecological Economics, 52, 273–288.

PRISM Group, 2008. http://www.prismclimate.org.

Pysek, P., Bacher, S., Chytry´ , M., Jarosˇik, V., Wild, J., Celesti-Grapow, L., Gasso, N., Kenis, M., Lambdon, P.W., Nentwig, W., Pergl, J., Roques, A., Sadlo, J., Solarz, W., Vila, M. & Hulme, P.E. (2010) Contrasting patterns in the invasions of European terrestrial and freshwater habitats by alien plants, insects and vertebrates. Global Ecology and Biogeography, 19, 317–331.

Regehr, D.L. (1988) Selective control of Japanese honeysuckle (*Lonicera japonica*). Weed Technology, 2, 139–143.

Reinhart, K.O., Greene, E. & Callaway, R.M. (2005) Effects of *Acer platanoides* invasion on understory plant communities and tree regeneration in the northern Rocky Mountains. Ecography, 28, 573–582.

Ricciardi, A. (2007) Are modern biological invasions an unprecedented form of global change? Conservation Biology, 21, 329–336.

Ricketts, T.H., Dinerstein, E., Olson, D.M., Loucks, C.J. & Eichbaum, W. (1999) Terrestrial Ecoregions of North America: A Conservation Assessment. Island Press, Washington, D.C.

Robertson, M.P., Villet, M.H. & Palmer, A.R. (2004) A fuzzy classification technique for predicting species' distributions: applications using invasive alien plants and indigenous insects. Diversity and Distributions, 10, 461–474.

Rose, N.A. & Burton, P.J. (2009) Using bioclimatic envelopes to identify temporal corridors in support of conservation planning in a changing climate. Forest Ecology and Management 258, S64–S74.

Ross, L.C., Lambdon, P.W. & Hulme, P.E. (2008) Disentangling the roles of climate, propagule pressure and land use on the current and potential elevational distribution of the invasive weed *Oxalis pes-caprae* L. on Crete. Perspectives in Plant Ecology, Evolution and Systematics, 10, 251–258.

SAS Institute (2009) SAS 9.2, Cary, NC.

Sasek, T.W. & Strain, B.R. (1990) Implications of atmospheric CO2 enrichment and climatic change for the geographical distribution of two introduced vines in the U.S.A. Climatic Change, 16, 31–51.

Schierenbeck, K.A. (2004) Japanese honeysuckle (*Lonicera japonica*) as an invasive species; history, ecology, and context. Critical Reviews in Plant Sciences, 23, 391–400.

Schweitzer, J.A. & Larson, K.C. (1999) Greater morphological plasticity of exotic honeysuckle species may make them better invaders than native species. Journal of the Torrey Botanical Society, 126, 15–23.

Smalley, G.W. (1979) Classification and evaluation of forest sites on the Southern Cumberland Plateau. Gen. Tech. Rep. SO-23. U.S. Department of Agriculture, Forest Service. Southern Forest Experiment Station, New Orleans, LA, 59 p.

Smalley, G.W. (1982). Classification and evaluation of forest sites on the Mid- Cumberland Plateau. Gen. Tech. Rep. SO-38. U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station, New Orleans, LA, 58 p.

Smalley, G.W. (1984) Classification and evaluation of forest sites in the Cumberland Mountains. Gen. Tech. Rep. SO-50. U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station, New Orleans, LA, 84 p.

Smalley, G.W. (1986) Classification and evaluation of forest sites on the Northern Cumberland Plateau. Gen. Tech. Rep. SO-60. U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station, New Orleans, LA, 74 p.

Standish, R.J., Robertson, A.W. & Williams, P.A. (2001) The impact of an invasive weed Tradescantia fluminensis on native forest regeneration. Journal of Applied Ecology, 38, 1253–1263.

Stohlgren, T.J., Binkley, D., Chong, G.W., Kalkhan, M.A., Schell, L.D., Bull, K.A., Otsuki, Y., Newman, G., Bashkin, M. & Son, Y. (1999) Exotic plant species invade hot spots of native plant diversity. Ecological Monographs, 69, 25–46.

Stohlgren, T.J., Chong, G.W., Schell, L.D., Rimar, K.A., Otsuki, Y., Lee, M., Kalkhan, M.A. & Villa, C.A. (2002) Assessing vulnerability to invasion by alien plant species at multiple spatial scales. Environmental Management, 29, 566–577.

Stohlgren, T.J., Ma, P., Kumar, S., Rocca, M., Morisette, J.T., Jarnevich, C.S. & Benson, N. (2010) Ensemble habitat mapping of invasive plant species. Risk Analysis, 30, 224–235.

Swenson, U., Stuessy, T.F., Baeza, M. & Crawford, D.J. (1997) New and historical plant introductions, and potential pests in the Juan Fernandez Islands, Chile. Pacific Science, 51, 223–253.

Timmins, S.M. & Williams, P.A. (1991) Weed numbers in New Zealand's forests and scrub reserves. New Zealand Journal of Ecology, 15, 153–162.

Truscott, A.M., Palmer, S.C.F., Soulsby, C. & Hulme, P.E. (2008) Assessing the vulnerability of riparian vegetation to invasion by *Mimulus guttatus*: relative importance of biotic and abiotic variables in determining species occurrence, abundance and performance. Diversity and Distributions, 14, 412–421.

Tucker, C.J. (1979) Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing and the Environment, 8, 127–150.

Underwood, E.C., Klinger, R. & Moore, P.E. (2004) Predicting patterns of alien plant invasions in Yosemite National Park, California, USA. Diversity and Distributions, 10, 447–459.

USBOC (2000) United States Bureau of the Census – Tiger Files, Geography Division.

USDA (2005) Forest Inventory and Analysis National Core Field Guide, Volume 1: Field Data Collection Procedures for Phase 2 Plots, Version 1.5. U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis, Washington Office. Internal Report.

USDA (2006) Land Resource Regions and Major Land Resource Areas of the United States, the Caribbean, and the Pacific Basin. U.S. Department of Agriculture Handbook 296. USDA, Washington, D.C.

USDA (2009) PLANTS National Plants Database. National Plant Data Center, Baton Rouge. http://plants.usda.gov.

USDA/FS (2007) The Forest Inventory and Analysis Database: Database Description and Users Guide Version 3.0. U.S. Department of Agriculture, Forest Service, Washington Office.

Vaclavik, T. & Meentemeyer, R.K. (2009) Invasive species distribution modeling (iSDM): are absence data and dispersal constraints needed to predict actual distributions? Ecological Modelling, 220, 3248–3258.

Vellend, M. (2002) A pest and an invader: white-tailed deer (*Odocoileus virginianus* Zimm.) as a seed dispersal agent for honeysuckle shrubs (*Lonicera* L.). Natural Areas Journal, 22, 230–234.

Vila, M., Basnou, C., Pyˇsek, P., Josefsson, M., Genovesi, P., Gollasch, S., Nentwig, W., Olenin, S., Roques, A., Roy, D., Hulme, P.E. & DAISIE partners (2010) How well do we understand the impacts of alien species on ecosystem services? A pan-European cross-taxa assessment. Frontiers in Ecology and the Environment, 8, 135–144.

Vitousek, P., D'Antonio, C., Loope, L., Rejmanek, M. & Westbrooks, R. (1997) Introduced species: a significant component of human-caused global change. New Zealand Journal of Ecology, 21, 1–16.

Walther, G.R., Roques, A., Hulme, P.E., Sykes, M.T., Pysek, P., Kuhn, I., Zobel, M., Bacher, S., Botta-Dukat, Z., Bugmann, H., Czucz, B., Dauber, J., Hickler, T., Jarosik, V., Kenis, M., Klotz, S., Minchin, D., Moora, M., Nentwig, W., Ott, J., Panov, V.E., Reineking, B., Robinet, C., Semenchenko, V., Solarz, W., Thuiller, W., Vila, M., Vohland, K. & Settele, J. (2009) Alien species in a warmer world: Risks and opportunities. Trends in Ecology and Evolution, 24, 686–693.

Wear, D.N. & Greis, J.G. (2002) Southern forest resource assessment: summary of findings. Journal of Forestry, 100, 6–14.

Webster, C.R., Jenkins, M.A. & Jose, S. (2006) Woody invaders and the challenges they pose to forest ecosystems in the Eastern United States. Journal of Forestry, 104, 366–374.

Williams, P.A., Timmins, S.M., Smith, J.M.B., Downey, P.O. (2001) The biology of Australian weeds 38, *Lonicera japonica* Thunb. Plant Protection Quarterly, 16, 90–100.

Willis, S.G. & Hulme, P.E. (2002) Does temperature limit the invasion of *Impatiens glandulifera* and *Heracleum mantegazzianum* in the UK? Functional Ecology, 16, 530–539.

With, K.A. & Crist, T.O. (1995) Critical thresholds in species' responses to landscape structure. Ecology, 76, 2446–2459.

Yates, E.D., Levia, D.F. & Williams, C.L. (2004) Recruitment of three non-native invasive plants into a fragmented forest in southern Illinois. Forest Ecology and Management, 190, 119–130.

CHAPTER 5.
Habitat Modelling of Alien Plant Species at Varying Levels of Occupancy

5.1 Abstract

Distribution modelling of invasive plants has broad application to planning effective land management, however it has limitations associated with modelling expanding populations in a changing environment and data availability. We assess predictive habitat models for three invasive plant species, at differing levels of occurrence using two different habitat-modelling techniques, logistic regression and maximum entropy. The influence of disturbance, spatial and temporal heterogeneity and other landscape characteristics are assessed by creating regional level models based on occurrence records from the United States Department of Agriculture, Forest Service, Forest Inventory and Analysis database. Logistic regression and maximum entropy (MaxEnt) models were assessed independently and evaluated as predictive tools to test the value of presence/absence and presence only data in predicting species distributions. Ensemble models were also developed that combined the predictions of the two modelling approaches to obtain a more robust prediction estimate. All species had strong models with AUC greater than 0.75. The species with the highest occurrence, Privet (*Lingustrum spp.*), had the greatest agreement between the models (93%) suggesting the relevance of the variables in developing this model. Tall fescue (*Lolium arundinaceum*), had the most disagreement between models at 33 percent and the lowest AUC values. This agrees with other studies that have found more difficulty predicting the distribution of shorter lived species. Overall, we showed the strength of integrative modelling in assessing and understanding habitat modelling.

5.2 Introduction

Predictive habitat modelling is a widely utilized tool in ecology (Guisan & Thuiller 2005; Smolik et al. 2010) that has broad applications in assessing relationships between species occurrence, the environment and the impact of ecological change (Guisan et al. 2006). For invasive species, habitat modelling is not only useful for predicting species distributions and ecological niches, but also for predicting, under differing future conditions, potential spread and

suitability of areas that have not yet been invaded. Predictive modelling can be used to assess the impacts of external environmental conditions such as climate change on species distribution (Kearney et al. 2010) and the potential impacts of the species itself on the landscape (Ficetola et al. 2002).

Invasive alien species are a major threat to ecosystems through the rapid anthropogenic acceleration of species introductions over the last century (Hulme et al. 2009) and the consequences on economies and ecosystems (Vilà et al. 2010). They are now recognised as a major aspect of global environmental change (Mainka & Howard 2010; Ricciardi 2007; Vitousek et al. 1997). As such, tools that can accurately assess the impacts of invasive alien species are essential to identify areas where management and monitoring efforts should be focused.

The strength of a habitat model is influenced by the correlation of species distribution to input parameters (Hoffman et al.2010) and the number of observation points. Input parameters are often landscape level digital information that provides a representation of environmental heterogeneity of the landscape, including: climate; habitat diversity; landscape characteristics; habitat patch size and shape; connectivity; regional and local diversity of biota; vegetation structure; and the intensity, frequency and magnitude of disturbance (Kumar et al. 2006; Stohlgren et al. 1999; With & Crist 1995), all of which vary across spatial and temporal scales (Kumar et al. 2006; Pickett & Cadenasso 1995). Collectively, these factors result in interlaced patterns of species distribution at multiple spatial and temporal scales (Wagner & Fortin 2005).

In this chapter, we develop predictive habitat models for three invasive plant species, privet (*Ligustrum spp*.), tall fescue (*Lolium arundinaceum*), silktree (*Albizia julibrissin*), that are at differing densities using two different habitat-modelling techniques, logistic regression and maximum entropy. We assess the impact of different landscape parameters, modelling techniques, and data set sizes on the final predictive models.

5.3 Methods

5.3.1 Study Area

The Cumberland Plateau and Mountain Region (CPMR) of United States extends from northern Alabama, through Tennessee, Kentucky and into Virginia (Smalley, 1979, 1982, 1984, 1986) (Figure 4-1). The CPMR covers a total area of 59,000 km^2, and has one of the most

diverse woody plant communities in eastern North America (Ricketts et al. 1999). Forest resources and management are a major part of the CPMR economy, particularly in rural communities. Approximately 70 percent of the land in this area is forested, with over 75 percent of this comprising hardwoods (Homer, 2004; USDA/FS, 2007). Elevations range from 200 m to 1200 m (Gesch et al. 2002), with annual rainfall varying from 940 mm to 1900 mm and mean minimum winter temperatures from -7 to 1.5 ^{o}C (PRISIM 2007).

Like many of the forests in eastern North America, the native deciduous hardwood forests of the CPMR are characterised by a long history of land-use change driven by agricultural conversion and timber extraction. More recently, urban sprawl and large-scale conversion of land to intensively managed pine plantations have become major contributors to land cover change (Wear & Greis 2001). McGrath and others (2004) found that 14 percent of native forest cover was lost between 1981 and 2000, predominantly as a result of native forest conversion to pine plantations. Of the 33 invasive species monitored by the USFS 25 are found on the CPMR: four trees, seven shrubs, seven vines, five grasses and two forbs.

5.3.2 Species of Interest

Study species were selected to represent a range of life forms and occurrence levels across the CPMR. One shrub (privet), one grass (tall fescue) and one tree (silktree) were selected to represent moderate occurrence (approximately 16% of sampled sites, privet), low occurrence (approximately 5% of sampled sites, tall fescue), and very low occurrence (approximately 2% of sampled sites, silktree) (Table 5-1).

Table 5-1: Number of points in training and test data sets for each species.

	Training		Test	
	occurrence	absence	occurrence	absence
Privet	200 (10.4%)	1125 (59.0%)	100 (5.2%)	482 (25.4%)
Tall fescue	65 (3.4%)	1270 (66.6%)	28 (1.5%)	544 (28.5%)
Silktree	31 (1.6%)	1304 (68.4%)	13 (0.7%)	559 (29.3%)

5.3.2.1 Privet

The USFS collects information on two species of privet Chinese privet (*L. sinense*) and European privet (*L. vulgare*) (USDA/FS 2007), however there are at least eight species of invasive privets that have been introduced from Asia and Europe into the southeastern United States as ornamentals (USDA 2011, Dirr 1998, Maddox et al. 2010). It can be difficult to distinguish between privet species thus we model the *Ligustrum* genus. Chinese privet is the most common invasive privet. It is a thicket forming shrub native to China and Europe. It was

introduced into the United States around the 1800s as a decorative shrub (Miller et al. 2010). It is now considered to be present in 20 states, ranging from Texas to Massachusetts (PLANTS 2011). Chinese privet is an evergreen shrub that can grow up to 10 m tall (Miller et al. 2010).

Privets are the second most abundant invasive plants in the southeastern region and most prevalent in the understory of bottomland hardwood forests (Merriam & Feil 2002, Miller et al. 2010). All species are still being produced, sold, and planted as ornamentals. Privets severely alter natural habitat and critical wetland processes forming dense stands to the exclusion of most native plants and replacement regeneration. The abundance of specialist birds and diversity of native plants and bees is dramatically reduced by privet thickets (Wilcox & Beck 2007, Hanula et al. 2009). The dense thickets impact forests communities by shading and out-competing many of the native species. Privet can survive in a variety of habitats, including wet or dry areas, but dominates best in mesic forests (Miller et al. 2010). Privets produce abundant seeds that are viable for about a year (Shelton & Cain 2002), which are predominately spread by birds (Greenberg & Walter 2010). They also have the ability to increase in density by stem and root sprouts. Controlling privet infestations costs the United States billions each year (Simberloff et al. 1997). The fruit produced does provide a substantial food source for birds and other wildlife. A study done on white tailed deer showed that privet is a key component of their autumn and winter diet and may also provide additional nutrients when mast production is scarce (Stromayer et al. 1998).

5.3.2.2 Tall fescue

Tall fescue (*Lolium arundinaceum*) is a grass native to Europe and was first introduced into the United States in the early to mid-1800s. It has been widely planted for turf, forage and erosion control (Hannaway et al. 1999). Tall fescue occurs throughout the continental United States (PLANTS 2011) and has been reported as invasive in natural areas (Fleming & Wofford 2004). Tall fescue is still promoted by a variety of agricultural agencies; however, the USDA FS Southern Region has prohibited the use of endophytically enhanced tall fescue on Forest Service lands (Miller et al. 2010). Tall fescue is a cool season grass that invades native grasslands, savannas, woodlands and other high-light natural habitats (Hannaway et al. 1999). It spreads mainly through rhizomes and can form extensive colonies that compete with and displace native vegetation. Viable seeds can be dispersed by grazing animals and birds and remain in the seed bank for extended periods of time (Miller et al. 2010).

Some varieties of tall fescue have a mutualistic fungal endophyte (*Neotyphodium coenophialum*) that gives them a competitive advantage over some plants, including legumes (Pedersen et al. 1990). As a result, communities dominated by tall fescue are often low in plant species richness (Spyreas et al. 2001). In addition, alkaloids produced by endophyte-infected tall fescue may be toxic to small mammals and of low palatability to ungulates (Clay & Schardl 2002). Tall fescue, which has replaced many acres of native grass, does not supply the type of food and cover that many birds need in order to thrive (Schardl et al. 2004). For example, tall fescue only supports a limited number of insects (Rudgers & Clay 2008), which are an important food for both quail and turkey. Grasslands dominated by endophyte-infected tall fescue are expected to support less total herbivore biomass, which in turn should support less predator biomass (Rudgers & Clay 2008; Schardl et al. 2004).

Tall fescue tolerates nutrient-poor and compacted soils and grows well in disturbed areas such as highway and railroad right of ways. Annual nitrogen inputs are needed to maintain optimal grazing conditions (Hannaway et al. 1999). Tall fescue is adapted to cool, humid climates with moist soils of a pH 5.5 to 7.0 (Hannaway et al. 1999). It will produce top growth when soil temperatures are as low as 5^{o}C so it continues growing into late autumn in the southeastern United States (Hannaway et al. 1999).

5.3.2.3 Silktree

Silktree is a legume native to south and eastern Asia. It is a small to medium-sized tree that can grow to 11 m tall. It was introduced to the United States in 1745 and widely planted as ornamental. Silktree is now found throughout the southeastern United States along roadsides, beside parking lots bordering power lines and encroaching into forests. Silktree reproduces both vegetatively and by seed (Miller et al. 2010). The seeds are encased with impermeable seed coats that allow them to remain dormant for many years (Creager 1992). Silktree is shade-intolerant, can grow in a variety of soils, produces large seed crops and sprouts when damaged. These attributes make it a strong competitor in open areas and forest edges. Dense stands of silktree severely reduce the sunlight and nutrients available for other plants (Miller et al. 2010). Although it can tolerate partial shade but is rarely found in forests with full canopy cover or at higher elevations (above 900 m) where cold hardiness is a limiting factor. However, it can become a serious problem along riparian areas where it becomes established along scoured shores and where its seeds are easily transported in water (Miller et al. 2010).

Although it has been identified as invasive to forest in the southeastern United States (Miller et al. 2010), is still being encouraged as a tree crop species (Ares et al. 2009). Ares and others (2009) state that in the southeastern United States silktree has been considered in agroforestry practices as goat (Addlestone et al. 1998) and cattle browse (Bransby et al. 1992), and for soil fertility improvement in permaculture systems (Matta-Machado & Jordan 1995; Rhoades et al. 1997; Jordan 2004). However, planting of silktree should be evaluated on a site-specific basis because it can become invasive especially in riparian areas (Loewenstein & Loewenstein 2005). This mixed message may increase the planting and thus the invasion potential of silktree in the next decade.

5.3.3 Invasive Plant Occurrence

The Forest Service Forest Inventory and Analysis (FIA) program of the United States Department of Agriculture (USDA) collects, analyses and reports information on the status, trends and conditions of forests within the United States. Periodic surveys of all forested land in the United States have occurred since 1928 (Birdsey & Schreuder 1992). Recent inventories have typically been conducted every five to seven years in the southeastern states, with approximately 20 percent of the points assessed every year (USDA/FS 2007). In the CPMR there are 2,814 FIA sites (USDA/FS 2007).

An extension of the FIA database focuses on invasive plants and assesses the occurrence and density of invasive species of concern. The FIA database was made available for our study as of 2008 with data from the last completed inventory cycle (2000 – 2005).

5.3.4 Landscape Variables

Landscape variables were categorised into six groups: Landsat, anthropogenic, environmental, climate, land use and water. Using ArcGIS (ESRI 2009) and ERDAS (ERDAS Inc. 2007), all variables were extracted from available digital information including Landsat imagery, classified land use data, roads, river, human population census and climate information. All variables were converted to 30 m x 30 m cells across the CPMR. These are described in detail in Chapter 4.

The total number of variables was 41 (Table 4-1). This initial set was reduced using exploratory data analysis to remove variables that were highly correlated this was done using Pearson's correlation. For variables with a high correlation ($r > 0.8$) only one variable was selected for input into further models. All input variables were mapped; NDVI75 and NDVI90-

75 showed strong blocks based on scenes, an artefact of instrumentation, and thus were not suitable for use in further analysis. This left a set of 28 variables (Table 4-1).

Descriptive statistics for the 28 variables were calculated for both the FIA and forest CPMR data sets to determine if FIA data can be extrapolated to the entire CPMR (Table 4-1). The FIA points had a mean within one standard deviation of the mean for the forested area of the CPMR for all variables. All but two variables had means within 0.2 standard deviations. The mean for the CPMR for farming within 500 m was within 0.56 standard deviations of the FIA data and forest within 100 m was within 0.37 standard deviations. In both cases the maximum and minimum were very similar suggesting that although there was some variation in the means they still represent the full range of the CPMR. Overall this shows that FIA data is a good representation of the CPMR.

5.3.5 Models

Two modelling techniques were used, logistic regression (Hosmer & Lemeshow 2000) and maximum entropy (Phillips et al. 2006). We also integrated information from both thelogistic regression and MaxEnt models using an ensemble approach. An important difference between the techniques is that logistic regression uses information on both occurrence and absence to estimate a predictive linear model whereas maximum entropy (MaxEnt) uses information from occurrence only. See Chapter 4 for a comparison of these modelling approaches for Japanese Honeysuckle.

The distribution of each species was modelled, following the methods of Chapter 4, using each group of variables (Landsat, anthropogenic, environmental, land use, water and climate) separately (Table 4-1). These “sub-models” were built using each of the two techniques, logistic regression and maximum entropy. Using only variables selected in the final sub-model for each variable group, a final composite model was determined. Logistic regression models were conducted using SAS (SAS Institute 2009) and maximum entropy models were conducted using a specialised package of MaxEnt (Phillips et al. 2006). Logistic regression models were derived using a backward stepwise regression method with Akaike's Information Criterion (AIC) (Akaike 1974) as the selection criterion. The MaxEnt models were derived using a manual backward selection method and variables that had little or no impact on the model were removed. Impact on the model was assessed by percent contribution, the results of a jack-knife

test on gain and based on influence on area under the curve (AUC). A measure of variable contribution was calculated to identify the key variables determining each species occurrence.

The omission rate, Cohen's Kappa and AUC were used to assess the reliability and validity of the models. The omission rate is the false negative or the proportion of sites where the species was present but the model predicted absence. The Kappa statistic is the chance-corrected proportional agreement where possible values range from +1 (perfect agreement) to -1 (complete disagreement), via 0 (no agreement above that expected by chance). To calculate omission rate and Kappa statistic the predicted model values are converted to a binary value (predicted occurrence, 1, or predicted absence, 0). The threshold value for this binary conversion was set, for each species, as the value that maximized the sum of sensitivity and specificity (Manel et al. 2002). Area Under the receiver operator Curve (AUC) provides a single measure of model performance independent of any particular choice of threshold (Lobo et al. 2008).

Rasters were imported from MaxEnt into ArcGIS and the raster calculator was used to apply the logistic regression model. Initial maps were continuous rasters that were reclassified into binary rasters based on the cut-off values determined by maximizing the sum of sensitivity and specificity.

While the logistic regression and MaxEnt models may be considered individually to select the best overall model for particular datasets, methods to combine the two models have the potential to reduce the uncertainty associated with any one particular algorithm. A number of approaches have been proposed for combining outputs of individual models for ensemble predictions (Araújo & New 2007). Here we adopt a consensus approach, identifying areas of agreement and disagreement in the models.

5.3.6 Data Selection

Models were built using 70 percent of the data with the remaining 30 percent used to test the models (Table 5-1). Logistic regression absence data were down sampled to give a 2:8 ratio with presence data (Oommen et al. 2010).

5.4 Results and Discussion

Of the 42 models run, 41 have better than random predictions (Table 5-2). All three species had low omission rates and high AUC. The final logistic regression and MaxEnt models were constructed as a composite of the most useful variables from the subgroup models. Then these

final logistic regression and MaxEnt models were combined to produce an ensemble model. These showed very strong agreement between the privet (highest occurrence) composite models with only 7 percent disagreement (Figure 5-1). This general trend of increasing consistency with other models is what would be expected, given more information the more constant the model. However, despite low prevalence, particularly for silktree, which had only 31 data points in the training set, models for all species were acceptable and should prove useful. The weakest model was that for tall fescue (test AUC of 0.75).

Of the 28 original variables used in development of these models, 15 were ultimately incorporated into a final composite models; only seven were used in more than one model (Table 5-3). Overall the composite models were dominated by environmental variables (32% of all composite model contributions) and climate variables (42% of all composite model contributions) with minimum temperature the single most important variable (40% of all composite model contributions) (Table 5-3). This confirms the utility of matching the native ranges of species with the range of potential invasion, and the approach of integrating elevation and latitude as is used to estimate potential invasive distribution (Peterson et al. 2003). It also suggests that climate change will influence the distribution of invasive species. Variables in the Landsat and water groups contributed very little to the models, only one variable each to the composite models (disturbance index in 2001 at one percent, and water within 500 m at three percent, of all composite model contributions) (Table 5-3). Information on human population, roads and land use (proportion of forest and proportion of farming) were the most useful anthropogenic variables (Table 5-3). All of this information is readily available for North America and much of the world, making this level of landscape level modelling very practical.

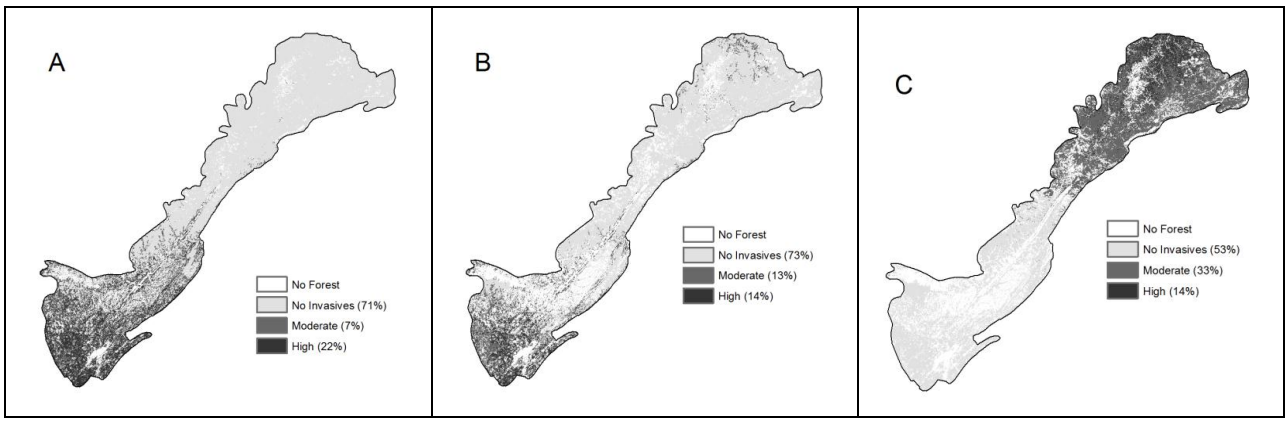

Figure 5-1: Spatial representation of combine composite models (A - privet, B - silktree, C - tall fescue)

Table 5-2: Threshold (defined as maximum sensitivity plus specificity) and accuracy assessment for the three species (bold denotes strong models, AUC>0.80 and omission rate <0.20) using logistic regression (L) and MaxEnt (M). The variables were grouped into four groups, Landsat, Anthropogenic, Environmental and Climate. The composite model is the final, best model.

Species	Model	Group	Threshold	Omission Rate		Kappa Coefficient		AUC	
				train	test	train	test	train	test
Privet	L	Landsat	0.18	0.07	0.10	0.05	0.03	0.70	0.66
	M	Landsat	0.47	0.25	0.37	0.08	0.05	0.74	0.68
	L	Anthro	0.16	0.06	0.08	0.14	0.07	0.76	0.72
	M	Anthro	0.33	0.09	0.20	0.10	0.09	0.77	0.70
	L	Enviro	0.16	0.04	0.05	0.04	0.01	0.84	0.82
	M	Enviro	0.34	0.10	0.10	0.05	0.06	0.80	0.80
	L	Climate	0.16	0.20	0.25	0.10	0.12	0.83	0.82
	M	Climate	0.42	0.18	0.30	0.09	0.11	0.83	0.82
	L	Land use	0.13	0.05	0.11	0.05	0.06	0.83	0.82
	M	Land use	0.30	0.10	0.13	0.04	0.03	0.81	0.79
	L	Water	0.17	0.10	0.21	0.01	0	0.66	0.65
	M	Water	0.52	0.51	0.51	0.02	0.05	0.66	0.67
	L	Composite	0.17	0.02	0.05	0.04	0.02	0.91	0.89
	M	Composite	0.28	0.07	0.14	0.03	0.04	0.86	0.83
Tall Fescue	L	Landsat	0.06	0.32	0.54	0.08	0.04	0.74	0.65
	M	Landsat	0.45	0.34	0.64	0.11	0.03	0.76	0.61
	L	Anthro	0.05	0.41	0.40	0.07	0.05	0.65	0.63
	M	Anthro	0.40	0.26	0.32	0.06	0.06	0.70	0.67
	L	Enviro	0.05	0.49	0.38	0.03	0.03	0.60	0.62
	M	Enviro	0.49	0.34	0.50	0.12	0.10	0.75	0.66
	L	Climate	0.06	0.28	0.30	0.10	0.08	0.73	0.70
	M	Climate	0.44	0.15	0.10	0.13	0.12	0.77	0.84
	L	Land use	0.06	0.36	0.42	0.05	0.04	0.66	0.59
	M	Land use	0.46	0.24	0.39	0.10	0.05	0.72	0.60
	L	Water	No Model						
	M	Water	0.47	0.20	0.32	0.03	0.01	0.61	0.54
	L	Composite	0.05	0.25	0.21	0.10	0.08	0.78	0.75
	M	Composite	0.42	0.25	0.35	0.19	0.14	0.82	0.75
Silktree	L	Landsat	0.02	0.41	0.38	0.02	0.03	0.69	0.60
	M	Landsat	0.47	0.29	0.36	0.06	0.05	0.73	0.73
	L	Anthro	0.02	0.22	0.41	0.04	0.03	0.75	0.73
	M	Anthro	0.47	0.29	0.21	0.16	0.16	0.84	0.90
	L	Enviro	0.02	0.10	0.15	0.07	0.05	0.83	0.80
	M	Enviro	0.30	0.06	0.29	0.07	0.04	0.82	0.78
	L	Climate	0.02	0.19	0.17	0.05	0.04	0.77	0.75
	M	Climate	0.42	0.13	0.07	0.05	0.05	0.77	0.82
	L	Land use	0.02	0.35	0.37	0.05	0.06	0.81	0.85
	M	Land use	0.42	0.25	0.43	0.07	0.05	0.80	0.70
	L	Water	0.02	0.29	0.35	0.05	0.04	0.74	0.75
	M	Water	0.49	0.32	0.28	0.03	0.05	0.76	0.76
	L	Composite	0.02	0.16	0.38	0.10	0.07	0.89	0.80
	M	Composite	0.27	0.06	0.07	0.14	0.13	0.91	0.90

Table 5-3: Contribution of variables to the final models (- negative, + positive, ∩ or U for binomial relationship), dominant variables given in bold.

	Species	Privet		Tall Fescue		Silktree	
	Model	L	M	L	M	L	M
Landsat	DI00					(+)6	
Anthropogenic	CENSUS						(+)24
	RD DEN	(+)4	(+)15			(+)35	(+)16
	RD DIST	(-)3					
	MRD DIST	(-)3					
	RES100			(-)8			
Environmental	DEM	(-)13	(-)7		(∩)19	(-)58	(-)48
	NORTHNESS			(-)30	(-)7		
	SLOPE		(-)6				
Climate	MINT	(+)66	(+)55	(-)62	(-)54		
	RANN		(U)5		(∩)10		
Land use	F00 100		(-)10				
	FARM500	(+)4			(∩)10		
	LULC90	7					
Water	WATER500					(-)1	(+)12
Proportion forest area invaded		24%	28%	46%	16%	20%	21%

Privet composite models used a range of environmental and anthropogenic variables, with the logistic regression model having seven variables and the MaxEnt model having six variables. The logistic regression model predicted 24 percent of the forest potentially invasible, and the MaxEnt model predicted 28 percent, while currently 15 percent of FIA plots have privet. Overall 22 percent of the area was predicted by both models. Both composite models were strong, with the logistic regression model slightly better. Environmental variables dominated both models, at 73 percent (MaxEnt) and 79 percent (logistic regression). Minimum temperature was the single most dominant variable, with higher minimum temperatures resulting in higher probability of occurrence. Both models suggest privet will be found at lower elevation in areas of higher road density (increased human occupation). The logistic regression model also suggested privet had a higher chance of occurrence closer to roads and with more farming in the near vicinity, while the MaxEnt model suggested that the less forest the more likely the area to have privet. The logistic regression model used historical land use as one of the independent variables, associating privet

to areas with less forest, more residential and more water in 1990. Overall this suggests that area of higher human use and disturbance would have more privet.

For tall fescue, the MaxEnt composite model had the highest kappa and highest AUC (Table 5-2), however, the MaxEnt model that used only climate variables had a slightly better omission rate. The logistic regression models had slightly lower validation statistics. Both the MaxEnt and logistic regression composite models were dominated by climate variables. The MaxEnt composite model showed that tall fescue occurrence was most influenced by temperature, elevation, rainfall, farming and aspect. Lower temperature, intermediate levels of farming, rainfall and elevation on more southerly aspects are related to higher occurrence of tall fescue. The logistic regression composite model only used three variables, minimum temperature, aspect and amount of residential in 100 m, with low temperature, more southerly slopes and less residential having higher occurrence of tall fescue.

The silktree was the only species for which anthropogenic variables were prominent in the composite models (Table 5-2). The MaxEnt composite model predicted 21 percent of the area to have probable occurrence of silktree and showed its occurrence to be influenced by elevation, population density, road density and water bodies. The variables lower elevation, higher population and road density and nearby water bodies related to the higher occurrence of silktree. The composite logistic regression model also used a number of anthropogenic variables, the model being dominated by elevation. However road density also had a major role in the model. The logistic regression composite model was the only composite model to use a Landsat variable. The logistic regression model also suggested that low elevation and high road density are important contributors to silktree occurrence. Higher disturbance in the landscape was also important, with distance from water weakly important.

Remote sensing has been identified as an emerging tool for biodiversity science and conservation (Turner et al. 2003). However, in this work, introduction of remotely sensed medium resolution (30 m) data had little value in the overall model development, with the Landsat models being not much better than random. Only one of the composite models, the logistic regression model for silktree, used any Landsat variables. The silktree model used the disturbance index for 2001 but it only added 5% to the model. Given the time put into developing the Landsat variables we would suggest that for future work this information adds little value to the predictive ability of models and is probably unnecessary at a landscape scale. The large size of the study area (59,000 km^2) made it impractical to use finer resolution remotely

sensed data due to the computer processing power required for analysis. However, exploration of different abstraction resolutions, as suggested by Sester (2005), would be worth exploring on a smaller scale to identify an optimal resolution.

The use of the two different modelling approaches, logistic regression and MaxEnt, strengthens the results. The inclusion in the models of similar variables with the same direction of relationships gives confidence to any inference about the importance of these variables. In examining all the composite models there was only one variable that had a different relationship between the two types of modelling, water, for the tall fescue composite models. In this model, water had a positive relationship with the MaxEnt model (12% contribution) but a small weak relationship in the logistic regression model (1% contribution to the model). The ensemble approach and mapping the agreement and disagreement of composite models within each species showed privet to have very strong agreement (97%), silktree moderate agreement (87%) and tall fescue limited agreement (67%). This is a reflection on the model constancy, number of occurrence points and the applicability of the independent variables in predicting the species of interest. Tall fescue did not have the smallest number of occurrence points, but had the lowest agreement. This suggests that for habitat modelling of grasses, such as tall fescue, there may be better alternative independent variables that could be used, or indeed that this may not be an appropriate approach for modelling such species. Another explanation is that only forested landscapes were modelled rather than grasslands.

Models such as those developed by this research can be used as tools for landscape management, forest stand assessment or long term forest monitoring programs. One of the greatest benefits of large-scale GIS models is that they can outline the main characteristics of species distribution areas and be used to predict environmental favourability in regions where their distribution is less documented (Barbosa et al. 2009). They can also be integrated into forest management decision support systems (Ducheyne et al. 2006) and assist in developing long term management plans.

5.5 Literature Cited

Addlestone, B.J., Mueller, J.P. & Luginbuhl, P.M. (1998) The establishment and early growth of three leguminous tree species for use in silvopastoral systems in the southern USA. Agroforest Systems, 44, 253-265.

Akaike, H. (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19, 716-723.

Araujo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. Trends in Ecology and Evolution, 22, 42–47.

Ares, A., Burner, D.M. & Brauer, D.K. (2009) Soil phosphorus and water effects on growth, nutrient and carbohydrate concentrations, d13C, and nodulation of silktree (*Albizia julibrissin* Durz.) on a highly weathered soil. Agroforest Systems, 76, 317-325

Barbosa, A.M., Real, R. & Vargas, J.M. (2009) Transferability of environmental favourability models in geographic space: The case of the Iberian desman (*Galemys pyrenaicus*) in Portugal and Spain. Ecological Modelling, 220, 747-754.

Birdsey, R.A. & Schreuder, H.T. (1992) An overview of forest inventory and analysis estimation procedures in the eastern United States-with an emphasis on the components of change. USDA Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado. USA. 11 p.

Bransby, D.I., Sladden, S.E. & Aiken, G.E. (1992) Silktree as a forage plant: a preliminary evaluation. Proceeding Forage Grassland Conference, Georgetown, TX, USA. Am Forage Grassl Counc 1, 28–31

Clay, K., & Schardl, C. (2002) Evolutionary origins and ecological consequences of endophyte symbiosis with grasses. University Chicago Press. 99-127 p.

Creager, R.A. (1992) Seed Germination, physical and chemical control of Catclaw Mimosa (*Mimosa pigra var. pigra*). Weed Technology, 6, 884-891.

Dirr, M.A. (1998) Manual of woody landscape plants: their identification, ornamental characteristics, culture, propagation and uses. Champaign, IL: Stipes Publishing L.L.C. 1,187 p.

Ducheyne, E.I., De Wulf, R.R. & De Baets, B. (2006) A spatial approach to forest-management optimization: linking GIS and multiple objective genetic algorithms. International Journal of Geographical Information Science, 20, 917-928.

ERDAS Inc. (2008) Imagine 9.2, 5051 Peachtree Corners Circle, Norcross, GA.

ESRI. (2009) ArcGIS, Redlands, CA. Environmental Systems Research Institute.

Ficetola, G.F., Thuiller, W. & Miaud, C. (2002) Prediction and validation of the potential global distribution of a problematic non-native invasive species; the American bullfrog. Diversity and Distributions, 8, 49–56.

Fleming, C.A. & Wofford, B.E. (2004) The Vascular Flora of Fall Creek Falls State Park, Van Buren and Bledsoe Counties, Tennessee. Castanea, 69, 164-184.

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M. & Tyler, D. (2002) The national elevation dataset. Photogrammetric Engineering and Remote Sensing, 68, 5–11.

Greenberg, C.H. & Walter, S.T. (2010) Fleshy fruit removal and nutritional composition of winter-fruiting plants: a comparison of non-native invasive and native species. Natural Areas Journal, 30, 312-321.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. Ecology Letters, 8, 993-1009.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J.M.C., Aspinall, R. & Hastie. T. (2006) Making better biogeographical predictions of species' distributions. Journal of Applied Ecology, 43, 386-392.

Hannaway, D., Fransen, S., Cropper, J., Teel, M., Chaney, M., Griggs, T., Halse, R., Hart, J., Cheeke, P. & Hansen D. (1999) Tall fescue (*Festuca arundinacea* Schreb). A Pacific Northwest Ext. Publ. PWN 504. Oregon State Univ., Corvallis.

Hanula, J.L., Horn, S. & Taylor, J.W. (2009) Chinese privet (*Ligustrum sinense*) removal and its effect on native plant communities of riparian forests. Invasive Plant Science and Management, 2, 292-300.

Hoffman, J.D., Aguilar-Amuchastegui, N. & Tyre, A.J. (2010). Use of simulated data from a process-based habitat model to evaluate methods for predicting species occurrence. Ecography, 33, 656-666.

Homer, C., Huang, C., Yang, L. & Wylie, B. (2004) Development of a 2001 national land cover database for the United States. Engineering and Remote, 70, 829-840.

Hosmer, D.W. & Lemeshow, S. (2000) Applied Logistic Regression. Wiley-Interscience, New York.

Hulme, P.E., Pyšek, P., Nentwig, W. & Vila. M. (2009) Will threat of biological invasions unite the European Union? Science, 324, 40-41.

Jordan, C.F. (2004) Organic farming and agroforestry: alley cropping for mulch production for organic farms in southern United States. Agroforest Systems, 61-62,79-90.

Kearney, M.R., Wintle, B.A. & Porter, W.P. (2010) Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. Conservation Letters, 3, 203-213.

Kumar, S., Stohlgren, T.J. & Chong, G.W. (2006) Spatial heterogeneity influences native and alien plant species richness. Ecology, 87, 3186-99.

Lobo, J.M., Jiménez-Valverde, A. & Real. R. (2008) AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography, 17, 145-151.

Loewenstein, N.J. & Loewenstein, E.F. (2005) Alien plants in the understory of riparian forests across a land use gradient in the Southeast. Urban Ecosystems, 8, 79-91.

Lunetta, R.S. & Elvidge, C.D. (1998). Remote sensing change detection. Ann Arbor Press, MI.

Maddox, V., Byrd, Jr., J. & Serviss, B. (2010) Identification and control of invasive privets (*Ligustrum spp*.) in the middle southern United States. Invasive Plant Science and Management, 3, 482-488.

Mainka, S.A. & Howard, G.W. (2010) Climate change and invasive species: double jeopardy. Integrative Zoology, 5, 102-111.

Manel, S., Williams, H.C. & Ormerod. S.J. (2002) Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology, 38, 921-931.

Matta-Machado, R.P. & Jordan, C.F. (1995) Nutrient dynamics during the first three years of an alley cropping agroecosystem in southern USA. Agroforest Systems, 30, 351-362.

McGrath, D.A, Evans, J.P., Smith, C.K., Haskell, D.G., Pelkey, N.W., Gottfried, R.R., Brockett, C.D., Lane, M.D. & Williams, E.D. (2004) Mapping land-use change and monitoring the impacts of hardwood-to-pine conversion on the Southern Cumberland Plateau in Tennessee. Earth Interactions, 8, 1-24.

Merriam, R.W. & Feil, E. (2002) The potential impact of an introduced shrub on native plants diversity and forest regeneration. Biological Invasions, 4, 369-373.

Miller, J.H., Chambliss, E.B. & Loewenstein, N.J. (2010) A field guide for the identification of invasive plants in southern forests. Gen. Tech. Rep. SRS–119. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 126 p.

Oommen, T., Baise, L.G. & Vogel, R.M. (2010) Sampling Bias and Class Imbalance in Maximum-likelihood Logistic Regression. Mathematical Geosciences, 43, 99-120.

Pedersen, J.F., Lacefield, G.D. & Ball, D.M. (1990) A review of the agronomic characteristics of endophyte-free and endophyte-infected tall fescue. Appl. Agric. Res., 3, 188-194.

Peterson, A.T. (2003) Predicting the geography of species' invasions via ecological niche modelling. Quart. Rev. Bio., 78, 419-433.

Phillips, S., Anderson, R. & Schapire R. (2006) Maximum entropy modelling of species geographic distributions. Ecological Modelling, 190, 231-259.

Pickett, S.T.A. & Cadenasso, M.L. (1995) Landscape ecology: spatial heterogeneity in ecological systems. Science, 269, 331-334

PRISM (2007) http://www.prismclimate.org. PRISM Group, Oregon State University.

Rhoades C.C., Nissen T.M. & Kettler J.S. (1997) Soil nitrogen dynamics in alley cropping and no-till systems on ultisols of the Georgia Piedmont, USA. Agrofor Syst 39, 31-44.

Ricciardi, A. (2007) Are modern biological invasions an unprecedented form of global change? Conservation Biology, 21, 329-36.

Ricketts, T.H., Dinerstein, E., Olson, D.M., Loucks, C.J. & Eichbaum, W. (1999) Terrestrial Ecoregions of North America: A Conservation Assessment. Washington, D.C., Island Press.

Rudgers, J. & Clay, K. (2008) An invasive plant–fungal mutualism reduces arthropod diversity. Ecology Letters, 11, 831-840.

SAS Institute (2009) SAS 9.2, Cary, NC.

Schardl, C.L., Leuchtmann, A. & Spiering, M.J. (2004) Symbioses of grasses with seedborne fungal endophytes. Annu. Rev. Plant Biol., 55, 315-340

Sester, M. (2005) Optimization approaches for generalization and data abstraction. International Journal of Geographical Information Science, 19, 871-897.

Shelton, M.G. & Cain, M.D. (2002) Potential carry-over of seeds from 11 common shrub and vine competitors of loblolly and shortleaf pines. Canadian Journal of Forest Research, 32, 412-419.

Smalley, G.W. (1979) Classification and evaluation of forest sites on the Southern Cumberland Plateau. Gen. Tech. Rep. SO- 23. New Orleans, LA: U.S. Department of Agriculture, Forest Service. Southern Forest Experiment Station. 59 p.

Smalley, G.W. (1982) Classification and evaluation of forest sites on the Mid-Cumberland Plateau. Gen. Tech. Rep. SO-38. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 58 p.

Smalley, G.W. (1984) Classification and evaluation of forest sites in the Cumberland Mountains. Gen. Tech. Rep. SO-50. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 84 p.

Smalley, G.W. (1986) Classification and evaluation of forest sites on the Northern Cumberland Plateau. Gen. Tech. Rep. SO- 60. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 74 p.

Smolik, M.G., Dullinger, S., Essl, F., Kleinbauer, I., Leitner, M., Peterseil, J., Stadler, L.M. & Vogl, G. (2010) Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. Journal of Biogeography, 37, 411-422.

Spyreas, G., Gibson, D. J. & Middleton B.A. (2001) Effects of endophyte infection in tall fescue (*Festuca arundinacea*: Poaceae) on community diversity. International Journal of Plant Sciences, 162, 1237-1245.

Stohlgren, T.J., Binkley, D., Chong, G.W., Kalkhan, M.A., Schell, L.D., Bull, K.A., Otsuki, Y., Newman, G., Bashkin, M. & Son, Y. (1999) Exotic Plant Species Invade Hot Spots of Native Plant Diversity. Ecological Monographs, 69, 25-46.

Stromayer, K., Warren, R.J., Johnson, A.S., Hale, P.E., Rogers, C.L. & Tucker, C.L. (1998) Chinese Privet and the Feeding Ecology of White-Tailed Deer: The Role of an Exotic Plant. The Journal of Wildlife Management, 62, 1321-1329

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E. & Steininger, M. (2003) Remote sensing for biodiversity science and conservation. Trends Ecol. Evol., 18, 306-314.

USDA (2011) PLANTS National Plants Database, http://plants.usda.gov. National Plant Data Center, Baton Rouge.

USDA/FS (2007) The Forest Inventory and Analysis Database: Database Description and Users Guide Version 3.0.

Vilà, M., Basnou, C., Pyšek, P., Josefsson, M., Genovesi, P., Gollasch, S., Nentwig, W., Olenin, S., Roques, A., Roy, D. & Hulme, P.E. (2010) How well do we understand the impacts of alien species on ecosystem services? A pan-European, cross-taxa assessment. Frontiers in Ecology and the Environment, 8, 135-144.

Vitousek, P.M., D'Antonio, C.M., Loope, L.L., Rejmanek, M. & Westbrooks, R. (1997) Introduced species: a significant component of human-caused global change. New Zealand Journal of Ecology, 21, 1-16.

Wagner, H.H. & Fortin, M.J. (2005) Spatial analysis of landscapes: concepts and statistics. Ecology, 86, 1975-1987.

Wear, D.N. & Greis, J.G. (2002) Southern forest resource assessment: summary of findings. Journal of Forestry, 100, 6–14.

Wilcox, J. & Beck, C.W. (2007) Effects of *Ligustrum sinense* Lour. (Chinese privet) on abundance and diversity of songbirds and native plants in a southeastern nature preserve. Southeastern Naturalist, 6, 535–550.

With, K.A. & Crist. T.O. (1995) Critical thresholds in species' responses to landscape structure. Ecology, 76, 2446-2459

CHAPTER 6.

Invasive Potential of Five Alien Trees in the Forest of the Southeastern Region, United States.

Collaborators: John Coulston and James H. Miller

6.1 Abstract

Alien trees introduced for commercial or landscaping use have caused substantial problems as invaders of natural and managed ecosystems. The magnitude of the problem has increased significantly over the past few decades, with accelerated land disturbance, land use changes, and global and internal transportation. In the southeastern region of the United States invasive plants are one of the threats to the long term sustainability of our forest ecosystems, along with climate and land use change. We assessed the potential distribution of five alien trees on forested lands in the southeastern region on forested lands. Maximum entropy algorithms were used to integrate anthropogenic and environmental variables with species occurrence data from the USDA Forest Service. All five species were predicted to significantly increase their distribution at the regional level under current conditions. Four climate change scenarios were then applied to the current potential. The five species responded very differently to the future scenario, particularly when climate variables had higher variability between current conditions and future scenarios. The variation between climate change scenarios with one species is greater than the variation between all species under current conditions. This suggests the importance of adaptive management and long term monitoring programs and the need for further development for assessing probable future climate conditions.

6.2 Introduction

Invasive species are those that establish and spread when introduced outside of their native range (Pysek 2004). Although few introduced species become naturalized and even fewer become invasive, those that do, often have undesirable ecological, aesthetic, and economic effects (Pimentel et al. 2005; Rejmanek et al. 2002) may modify ecosystem processes and affect ecosystem structure and functioning (Mack & D'Antonio 1998; Wilcove et al. 2000) with

economic impacts reaching billions of dollars (Sharma et al. 2005). They are an increasing management priority for governments worldwide owing to their potential to cause severe ecological and economic impacts (Ricciardi et al. 2011). The ecological effects of invasive species can range from modifications in ecosystem function and community structure to the extirpation or extinction of native species (Evans et al. 2011). At the global scale, the supply of many ecosystem services has declined over the last fifty years partially due to the spread of alien plants. At a regional scale, the southeastern forests of the United States are experiencing significant invasion of alien plants (Martin et al. 2009; Miller et al. 2012). The impact alien species are having may be exacerbated with climate change.

Climate changes are likely to affect the distributional ranges of native and invasive species (Elith et al. 2010; Hellmann et al. 2008; Walther et al. 2009). Due to the increased access to data, climate based distribution models are increasingly being used to explore future scenarios (Jeschke & Strayer 2008; Peterson 2003). The availability of these data and the use of more appropriate techniques allow for more accurate prediction of ranges and increase the ability to assess further assess the influence of climate change. It has been suggested that climate-based distribution models can predict areas at risk for invasion and explain the minimum requirements for a species to become established (Herborg, et al. 2007; Nuñez & Medley 2011; Zambrano & Martínez-Meyer 2006). However, the transition for a species from introduction to invasion is influenced by other factors beyond climate. A region having a similar climate to the native region is not sufficient to predict success (Nuñez & Medley 2011). Distribution modelling that relates species distribution to biologically relevant environmental variables can be used to assess potential invasion at a global level based on climatic conditions, or at a regional level by integrating further environmental parameters.

Species distribution models (SDM) have become a valuable tool for forecasting biological invasions (Peterson & Vieglais 2001). Predictions are based on niche theory, which states that species distribution patterns are governed by a discrete set of ecological conditions delineating the areas in which a species can establish and maintain populations at particular densities (Elton 1927; Hutchinson 1987; Vandermeer 1972). This modelling approach relies on known locations of a species to serve as response/training data, and relevant environmental variables as explanatory/predictor variables. The choice of the modelling approach and explanatory data affect the model estimates. These data should be determined based on study goals (Anderson & Raza 2010). When applying a modelling approach to invasives species there is the confounding

factor that species records do not necessarily reflect stable relationships with environment. Also, future invasion might involve environmental combinations that have not been adequately sampled or defined (Menke et al. 2009). One technique is to generate an ensemble of predictions, either though the application of several different modelling methods and/or data sets (Thuiller 2007). The final prediction can then emphasize agreement of predictions and can assist in identifying any model anomalies.

In this paper we develop regional species distribution models for five invasive trees and extend this by integrating climate change scenarios. Overall we assess the relative influence of disturbance, spatial heterogeneity, topographical and climate characteristics on current and future distributions.

6.3 Methods

6.3.1 Study Region

This study focuses on invasion of non-native trees in the forest of the southeastern region of the United States. The southeastern region is defined as 13 states (AL, AR, FL, GA, KY, LA, MS, NC, east OK, SC, TN, east TX, and VA). The region has numerous climatic zones, including temperate, sub-tropical, tropical, and arid (Hijmans et al. 2005). The landscape is heavily influenced by human use with the majority of forests having been harvested at least once in the last two centuries (Wear & Greis 2002). However forests remain the dominant land cover of this region, at 40 percent (Smith et al. 2009). The most common forest communities are hardwood, pine, mixed, and those with swamp adapted tree species. This range of tree communities supports some of the most biologically diverse forest in the world (Griep & Collins 2012). Of the 380-plus recognized invasive plants in southeastern forests and grasslands (more than 330 terrestrials and 48 aquatics) 53 are ranked as high-to-medium risk to natural communities (USDA Forest Service 2008).

6.3.2 Species of Interest

We focus on the five most prevalent invasive tree species invasive in the southeastern region: tree of heaven (*Ailanthus altissima*, [Mill.] Swingle), chinaberry tree (*Melia azedarach*, L.), silktree (*Albizia julibrissin*, Durazz.), princesstree (*Paulownia tomentosa*, [Thunb.] Siebold & Zucc. Ex Steud.) and tallowtree (*Triadica sebifera*, [L.] Small).

6.3.2.1 Tree of Heaven

The native range of tree of heaven covers large parts of China, where the species grows at low abundance as a natural component of broadleaf forests (Kowarik & Saumel 2007). Tree of heaven was first imported to the United States in the late 18th century and was widely used as an urban ornamental (Miller et al. 2010). It is now naturalized throughout a large portion of the continental United States, and is reported as established in the wild and has become a potential problem in forty-two states (USDA 2011). Tree of heaven is also naturalized in all other continents with the exception of Antarctica (Kowarik & Saumel 2007; Webb et al. 1988; Weber 2003). It is a dioecious tree that can reproduce vegetatively using root sprouts or through seed (Miller et al. 2010). Rapid establishment and first year growth under harsh conditions makes it a successful early colonizer, especially in disturbed sites (Knapp & Canham 2000; Miller 1990). Leaves and roots have allelopathic chemicals that inhibit the growth of other plants, thereby decreasing local biodiversity (Heisey 1996; Lawrence et al. 1991). Although tree of heaven is associated primarily with urban environments (Kowarik & Saumel 2007), it is also found in second- and old-growth forests (Carter 2007; Knapp & Canham 2000). Transportation corridors tend to promote the spread of tree of heaven from urban areas (Huebner 2003). Invasion often begins along forest roads where it spreads into recently harvested or disturbed sites, influencing stand development through exceptional competitive capabilities (Huebner 2003; Miller 1990).

Tree of heaven can tolerate a wide range of climatic conditions. Saplings and seedlings acclimate to heat, but they are negatively affected by cold and frost in comparison to other co-occurring plants (Kowarik & Saumel 2007). It acclimates well to droughts, and with elevated temperatures it switches biomass allocation from leaf production and stem elongation to root growth and stem diameter increase, increasing the uptake in moisture (Kowarik & Saumel 2007). Growth is best in nutrient rich, loamy soils but it can also tolerate nutrient-poor soils (Miller 1990).

6.3.2.2 Chinaberry Tree

Chinaberry tree is a fleshy-fruited deciduous tree from Asia (Pennington 1981) that grows to about 17 m. It was introduced to the United States in 1800s, and is now considered an escaped ornamental (Miller et al. 2010). It is a weedy tree that typically reproduces by prolific seeding, and may have clonal growth following disturbance (Tourn et al. 1999). Chinaberry outcompetes native vegetation by forming dense thickets (Tourn et al. 1999), and can produce viable seeds by

four years. Its fruit is widely consumed by birds, which are the primary seed dispersers (Miller et al. 2010). Chinaberry grows best along forest boarders, as well as disturbed areas in the Southeast, but is uncommon in higher elevations. Currently chinaberry's range stretches coast to coast, from Florida to California, and as far north as Maine (USDA 2011).

6.3.2.3 Silktree

Silktree is a legume, native to south and east Asia that can grow up to 11 m. It was introduced to the United States in 1745, and was widely planted as an ornamental (Miller et al. 2010). It is now common throughout the southeastern United States along roadsides, parking lots, and power lines, and is also encroaching into forests. Silktree reproduces both vegetatively and by seed (Miller et al. 2010). The seeds have impermeable seed coats that allow them to remain dormant for many years (Creager 1992). Seeds are mostly dispersed below or around the parent plant, but can be widely dispersed by water. Silktree grows rapidly under good conditions but is short-lived and has weak, brittle wood. If cut or top-killed, trees re-sprout quickly, and sprouts can grow over one metre in a single season. It grows best in full sun locations and is not particular to soil type but has low salt tolerance. Because it can grow in a variety of soils, produce large seed crops, and re-sprout when damaged; it is a strong competitor to native trees and shrubs in open areas or forest edges. Dense stands of silktree severely reduce sunlight and nutrients available for other plants (Miller et al. 2010). It can tolerate partial shade but is rarely found in forests with full canopy cover, or at higher elevations (above 900 m), where cold hardiness is a limiting factor. It can, however, become a serious problem along riparian areas, where it becomes established along scoured shores and where its seeds are easily transported in water (Miller et al. 2010).

6.3.2.4 Princesstree

Native to East Asia, princesstree was introduced into the eastern United States in early 1800s (Miller et al. 2010), and is now found in 25 states in the east and south (USDA 2011). It is still widely sold and planted as an "instant" shade tree. Until recently, most research on princesstree in the United States is focused on increasing growth in plantations due to the exceptional timber value in exports to Japan (Johnson et al. 2003; Miller et al. 2010). In the northeast United States, princesstree plantations can produce valuable high quality wood, but in the southeastern region,

due to the more favourable growing season, tree growth is too fast, producing low-density wood that is of much lower quality and value.

The presence of princesstree is associated with natural disturbance (Williams 1993) and is therefore likely to be promoted by anthropogenic disturbance. Williams (1993) classified the species as a non-aggressive species, though others (Langdon & Johnson 1994) suggested that in areas of high disturbance it shows invasive traits. Although sun-adapted and capable of extremely rapid growth in high light environments, princesstree is tolerant of a wide range of light levels (Longbrake & McCarthy 2001). Forest management practices can influence the establishment and development of this species with growth and survival on clearcuts being greater than in forest edges or in undisturbed forest (Longbrake 2001).

The fruit is produced in capsules that open when the seeds are mature. Seeds are small, winged, and wind-dispersed (Miller et al. 2010), each capsule releasing several thousand seeds. Trees can produce viable seed in as little as five years, with wind dispersal up to three kilometres (Langdon & Johnson 1994). Germination requires bare soil, so it invades well after fire, harvesting, and other disturbances.

6.3.2.5 Tallowtree

Tallowtree is native to China and Japan. It was introduced into the United States in the 1700s in South Carolina (Hunt 1947) and later distributed in the Gulf Coast region in the 1900s by the United States Department of Agriculture (USDA) to establish a soap making industry. It is a deciduous tree, growing 20 m tall and forming nearly pure stands in former wetlands. It is more prevalent on low and flat lands, areas adjacent to water and roadways, sites recently harvested or disturbed, younger stands, and private forestlands (Gan et al. 2009). Tallowtree is shade intolerant, which limits seedling establishment in intact forests (Pattison & Mack 2009). It is often spread inland from coastal infestations by hurricanes (Chapman et al. 2008). Seeds are high in fat and protein and thus are regularly consumed and spread by birds and possibly mammals (Conway et al. 2002; Renne et al. 2002). Saplings as young as three years can produce viable seed and remain reproductive for up to 100 years, producing 100,000 seeds per year (Bruce et al. 1997). Infestations intensify by prolific surface root sprouts that have high tolerance to insect defoliation (Rogers & Siemann 2005) and have allelopathic properties (Conway et al. 2002). Increases in both range and severity of tallowtree invasions are predicted with warming climate

trends (Gan et al. 2009; Stohlgren et al. 2002). This invasive is dispersal limited, and has yet to occupy the full extent of its range in United States (Pattison & Mack 2009).

6.3.3 Invasive Plant Occurrence

Regional models were derived from USDA Forest Service, Forest Inventory Analysis (FIA) data. FIA collects, analyses, and reports information on the status, trends and conditions of forests within the United States. Periodic surveys of the nation's forested land have occurred since 1928 (Birdsey & Schreuder 1992). Currently an annual survey is conducted in a rotating panel design where in the southeastern region a systematic sample of 20 percent (5 year cycle) or 14.3 percent (7 year cycle) of the field plot locations are observed each year. The nominal sampling intensity is approximately 1 plot per 2403 ha of land area (Bechtold & Patterson 2005). An extension of the FIA database focuses on invasive plants and assesses the occurrence and density for species that are a concern. Presence/absence information for the five species was extracted from the invasive component of the FIA database. A species is recorded as present if one tree is found in the FIA plot (673 m^2). Thus in this paper invaded forest refers to at least one tree per 673 m^2. FIA data has been widely applied to species distribution modelling (Edwards et al. 2006; Lemke et al. 2011; Wang et al. 2011; Zimmermann et al. 2007) and the impact of climate change (Gan et al. 2009; Iverson et al. 1999; Iverson et al. 2007; Schwartz et al. 2001). The invasive plants are only assessed on forested lands thus the predictive models are only for forested lands of the southeastern region.

6.3.4 Landscape variables

Climate variables were derived from a global dataset that is easily accessible, Bioclim data (www.worldclim.org). Bioclim variables are derived from monthly temperature and precipitation values to generate biologically meaningful variables. They were downloaded at a 30-second resolution (approx 1 km x 1 km). Only variables that were determined to be uncorrelated were used: diurnal range (BIO2), minimum temperature in the coldest month (BIO6; min temp), mean temperature of wettest quarter (BIO8), mean temperature of driest quarter (BIO9), annual precipitation (BIO12; rain), precipitation in the driest month (BIO14), precipitation seasonality (measured as the coefficient of variation) (BIO15; rain CV) and precipitation of warmest quarter (BIO18; warm rain) (Table 6-1).

Climate change for 2060 was assessed using four downscaled global climate change model (GCM) data developed by the USFS (Wear & Greis 2012). Each GCM was spatially downscaled to 1/12 degree (5 arc minute) using ANUSPLIN, an interpolation model that incorporates four dimensions (climatic variable, latitude, longitude, and elevation) to produce gridded surfaces for both monthly precipitation and surface air temperature. The GCMs selected by Wear and Greis (2012) and used here were the MK2 (scenario A) and MK3.5 (scenario B) from the Australian Commonwealth Scientific and Industrial Research Organization (CSIRO), the HadCM3 (scenario C) from the United Kingdom Meteorological Centre, and the MIROC 3.2 (scenario D) from the Japanese National Institute for Environmental Studies. Forecasts from a variety of models consistently indicate a warmer future, with average annual temperatures increasing 2.5 to 3.5 °C by 2060. Precipitation forecasts are much more variable across the models, generally ranging between historical levels and levels that are somewhat lower with high spatial variability across the South. At the subregion level, there is a higher degree of uncertainty for some places like Florida and western Texas, but more consistency in others such as the drier conditions predicted for Arkansas and Oklahoma.

A total of 15 regional landscape variables were extracted from available digital information including classified land use data, roads, rivers, human population census, and elevation information using ArcGIS (Table 6-1). All variables were converted to 90 m x 90 m cells across the Southeastern region. The selection of these variables was based on previous work (Lemke et al. 2011). Four anthropogenic-related variables (distance to road, distance to main roads, distance to cities, and people km^{-2}) were derived from the 2000 Topologically Integrated Geographic Encoding and Referencing system. Roads, towns and census data were obtained from United States Census Bureau (USBOC 2000). Roads have been shown to influence both alien plant presence (Gelbard & Belnap 2003) and forest community structure (Avon et al. 2010). Human influence, both in original introduction effort and continual dispersal, can also influence the success and distribution of an invasive plant (Wilson et al. 2007), thus high density of people may relate to high probability of invasion. Two geophysical variables (slope and northness) were derived from a 90 m digital elevation model (DEM), part of the United States Geological Survey (USGS) National Elevation Dataset (Gesch et al. 2002). Seven land use variables were extracted from the USGS National Land Cover Data (NLCD) for 1992 and 2001 based on eight land uses (Anderson 1976). Proportion of forest (combined deciduous, evergreen, and mixed forest), farming, grassland, residential and pines in 2001 and the change in forest and

change in pines between 1992 and 2001 were calculated in a 500 m buffer using a moving window average. Streams may affect the distribution and establishment of plants by influencing seed dispersal and moisture availability. Riparian areas have been shown to contain more alien plants than nearby upland areas (Gan et al. 2009; Stohlgren et al. 2002). One variable, distance to stream, was derived from stream data.

Table 6-1: All regional landscape variables used in developing species distribution models.

Code	Variable	Units	FIA			Landscape		Source
			Mean	SD	Range	Mean	SD	
BIO2	Mean Diurnal Range	°C x10	130	7	79 to 149	128	9	Bioclim
Min Temp	Min Temperature of Coldest Month	°C x10	-5	35	-97 to 155	-2	40	Bioclim
BIO8	Mean Temperature of Wettest Quarter	°C x10	175	69	8 to 277	180	68	Bioclim
BIO9	Mean Temperature of Driest Quarter	°C x10	163	67	-32 to 284	166	67	Bioclim
Rain	Annual Precipitation	mm	1314	152	602 to 2067	1304	149	Bioclim
BIO14	Precipitation of Driest Month	mm	75	13	23 to 155	74	14	Bioclim
Rain CV	Precipitation Seasonality (C.V)	%	20	8	6 to 69	20	10	Bioclim
Warm Rain	Precipitation of Warmest Quarter	mm	350	76	183 to 244	350	83	Bioclim
DEM	Elevation	m	175	194	0 to 1752	222	213	Gesch et al. 2002
NORTH	Cosine of aspect multiplied by the sine of slope		0.01	0.40	-1 to 1	0.01	0.43	Piedallu & Gégout, 2008
SLOPE	Slope	degrees	2.2	3.9	0 to 62	1.7	3.4	ESRI, 2009
CENSUS	Number of people in 2000	km^{-2}	25	74	0 to 3036	26	78	USBOC 2000
RD	Distance of the FIA site to closest road	m	349	420	0 to 13110	336	364	USBOC 2000
MRD	Distance of the FIA site to closest main road	m	8693	7709	0 to 60616	8622	7607	USBOC 2000
CITY	Distance to cities in 2000	m	35459	20124	285 to 150092	36214	20869	USBOC 2000
FOR	Forest in 2001 within a 500 m buffer	proportion	0.57	0.30	0 to 1	0.7	0.23	Anderson et al. 1976
FARM	Farming in 2001 within a 500 m buffer	proportion	0.13	0.18	0 to 1	0.11	0.16	Anderson et al. 1976
GRASS	Grassland in 2001 within a 500 m buffer	proportion	0.05	0.09	0 to 0.92	0.04	0.07	Anderson et al. 1976
PINE	Pines in 2001 within a 500 m buffer	proportion	0.22	0.24	0 to 1	0.25	0.26	Anderson et al. 1976
RES	Residential in 2001 within a 500 m buffer	proportion	0.05	0.09	0 to 1	0.05	0.08	Anderson et al. 1976
FORC	Change in forest (1992 to 2001)	proportion	-0.10	0.21	-1 to 1	-0.06	0.18	Anderson et al. 1976
PINEC	Change in pines (1992 to 2001)	proportion	0	0.20	-1 to 1	0.03	0.2	Anderson et al. 1976
RIV	Distance to Rivers	m	273	501	0 to 11900	246	447	seamless.usgs.gov

6.3.5 Models

The maximum entropy algorithm (Phillips et al. 2006) was used to develop probability envelopes for the occurrence of the five invasive trees in the southeastern region. Maximum entropy is a machine learning technique that predicts species distributions using detailed geospatial data sets together with species occurrence information, conducted using a specialized package of MaxEnt (Phillips et al. 2006). It generally performs as well or better than other algorithms in tests of model performance (Elith et al. 2006; Phillips et al. 2006). MaxEnt can use correlated variables, but this can confound the comparison of variables used between models and substantially slow model processing. Pearson's correlation was used to assess the correlation between each pair of variables. If the correlation between two environmental variables was over 0.5 r^2 one of the variables was removed from the analysis. In selection of the uncorrelated variables, those that represent the extremes were selected (i.e. minimum temperature rather than mean temperature) as it has been suggested that these are more ecologically relevant to plant species (Zimmermann et al. 2009).

Models were derived using a manual backward selection method, in which variables that had little or no impact on the model were removed. Impact on the model was measured as percent contribution and with a jack-knife test on gain and influence on area under the curve (AUC). This allowed identification of key variables in determining the occurrence of each species. Models were run as ten random replicates for each species to determine variable contribution and undertake backward selection. Only one model run was used to assess test AUC, omission rates, and map species distribution.

Three techniques were used to assess model reliability; the use of test and training data, omission rate and AUC. Occurrence data were split with 30 percent in test and 70 percent in training datasets (Elith et al. 2006; Graham & Hijmans 2006). The omission rate is the type II error or the proportion of sites where the species was present but the model predicted absence. To calculate this, a cut-off criterion is required to convert continuous model predictions to binary classifications. We used a threshold value that maximized the sum of sensitivity and specificity. This has the advantage of giving equal weights to the probability of success of both presences and absences (Manel et al. 2002). This is one of the most appropriate methods to derive a binary variable from continuous probabilities when species presence-absence distribution data are unbalanced (Jiménez-Valverde & Lobo 2006; Liu et al. 2005). AUC provides a single measure of model performance, independent of any particular choice of threshold but is sensitive to the

method in which absences in the evaluation data are selected (Lobo et al. 2008). It is most applicable to data with true absences (Jiménez-Valverde 2011), thus it needs to be used with caution with datasets that don't have absence data. Species absences can be determined from FIA data. The absences are true absences, but may not be due to unsuitability of the site for the species, but rather to the failure of the species to reach the site.

We used the following classes of AUC to assess model performance: 0.50 to 0.75 = fair, 0.75 to 0.92 = good, 0.92 to 0.97 = very good, and 0.97 to 1.00 = excellent (Hosmer & Lemeshow 2000). In this application AUC is useful when used in conjunction with other validation statistics, as invasive species are often not at equilibrium with their environment and their current realized distribution is much smaller than their potential distribution, thus even field absence data can be temporary with a time dimension.

To estimate optimal range for each species the threshold value was determined by the maximized sum of sensitivity and specificity. Maps were created in ArcGIS. Initial maps were continuous raster's that were reclassified into binary raster's based on the threshold values determined by the maximized sum of sensitivity and specificity. Four climate change scenarios were applied to the models.

6.4 Results

No high correlations were identified among environmental variables, thus all variables were kept for further analysis (Table A9-13). The highest correlation was between elevation and minimum temperature of coldest month, this was less than r = 0.70, slightly lower than the threshold used to determine high correlation.

The analysis indicated all species had models that were better than random, with test AUC's ranging from 0.76 to 0.92 and test omission rates between 0.11 and 0.32 (Table 6-2). Of the 22 variables available for models, only ten were selected in the models. Minimum temperature of coldest month was the most dominant variable, it was used in all five models with an average contribution of 51 percent (Table 6-3). All species currently occupy less than three percent of the forest in the southeastern region and under current conditions have the potential to increase their ranges at least four fold (Table 6-4). Climate change scenario D was very different from the other scenarios with winter temperature being lower than the current winter temperature (Table 6-4). Due to the uncertainty in the potential climate change we adopted an ensemble approach where agreement in four or more models would suggest a very likely potential of the species to

invade the forest in the next 50 years (Table 6-5) and one or less would suggest the potential for invasion was very unlikely.

Table 6-2: Threshold (defined as maximum sensitivity plus specificity) and accuracy assessment for full models and proportion of forest with potential of invasion for five invasive trees of the Southeastern United States.

Species	n	Threshold	Omission Rate		AUC	
			train	test	train	test
Tree of Heaven	855	0.244	0.08	0.11	0.89	0.87
Chinaberry	232	0.278	0.11	0.16	0.88	0.87
Silktree	681	0.381	0.24	0.32	0.79	0.76
Princesstree	484	0.418	0.21	0.25	0.82	0.85
Tallowtree	1093	0.161	0.10	0.12	0.93	0.92

Table 6-3: Contribution of dominant variables to regional models of invasive five tree species, based on ten replicates (- negative, + positive, ∩ or U for quadratic relationship).

	Tree of Heaven	Chinaberry	Silktree	Princesstree	Tallowtree	Overall Contribution
Min Temp	-58	∩ 52	∩ 18	-64	∩ 63	51
Rain	-7		∩ 7			3
Rain CV	∩ 26		∩ 8	-15	∩ 23	14
Warm Rain		-13		∩ 11		5
DEM	∩ 9	∩ 9	∩ 18			7
Census			∩ 24			5
Roads			-33			6
Farm		∩ 18				3
Pine		-8		-9		3
Forest					-14	3

Table 6-4: Percentage of forest with potential for invasion by five alien tree species, under current conditions and four different climate change scenario (FIA give the current percent of FIA plots currently invaded).

	Tree of Heaven	Chinaberry	Silktree	Princesstree	Tallowtree
FIA	2	1	2	1	3
Current	33	22	17	33	12
Scenario A	13	48	17	16	36
Scenario B	24	39	25	27	28
Scenario C	3	57	20	2	64
Scenario D	35	<1	17	18	<1

Table 6-5: Spatial congruence (%) between species distribution models for five alien tree species under five different scenarios (current conditions and four potential climate change scenarios)

Number of Models	Tree of Heaven	Chinaberry	Silktree	Princesstree	Tallowtree
0	56	57	40	65	34
1	14	15	13	12	30
2	9	11	10	5	10
3	9	10	16	4	15
4	11	6	21	10	11
5	1	1	<1	5	<1

6.4.1 Tree of Heaven

Tree of heaven had a strong model with a test AUC of 0.87 and test omission rate of 0.11 (Table 6-2). The model identified the northern region as having high invasion probability (Figure 6-1). In comparison with the FIA county maps, there are a few points that have been missed in the prediction by the regional models. These areas are all at low densities (1 plot per county) and in the southern portion of the region. Minimum temperature of the coldest month was the most dominant variable, contributing 58 percent to the model. Tree of heaven was most likely to occur at low minimum temperature in coldest month (<-2°C) where there is also low annual precipitation (<1500 mm), with little seasonality in precipitation (8 to 18) and at intermediate elevations (150 to 950 m) (Table 6-3). Under all but scenario D, the potential distribution in 50 years is reduced (Table 6-4) with the distribution moving further north (Figure 6-1). To assess the potential future distribution, models were combined (Figure 6-6, Table 6-5). Seventy percent of the forest was predicted to have no tree of heaven under all the scenarios, with 12 percent predicted to have tree of heaven.

6.4.2 Chinaberry Tree

Chinaberry had a strong model with a test AUC of 0.87 and test omission rate of 0.16 (Table 6-2), the model identified the southern part of the region as having high invasion probability (Figure 6-2). The FIA distribution map coincides well with the regional probability map (Figure 6-2). Minimum temperature of coldest month was the dominant variable, contributing 52 percent to the model. Chinaberry was most likely to occur at moderate to high minimum temperature in coldest month (-1°C to 5°C), at low elevations but not right along the coast (20 to 200 m), where there is also low to moderate precipitation in the warmest quarter (<480 mm), and is more likely to occur in areas of higher farming (Table 6-3). Under most of the climate change scenarios the range of chinaberry is likely to increase, with scenario D the only one that has a reduction in distribution (Table 6-4). Seventy two percent of the forest was predicted to have no chinaberry, while it was predicted to occur at seven percent of forest sites (Figure 6-6, Table 6-5).

6.4.3 Silktree

Silktree had the weakest model with a test AUC of 0.76 and test omission rate of 0.32 (Table 6-2), the model identified most of the region as having some invasion probability (Figure 6-3). The mapped FIA county level data (Figure 6-3) appears like the modelled data, in that the

species is spread throughout the region with no clear geographical patterns. Minimum temperature of coldest month was still included in the model but did not dominate as it did with the other species. There were more variables used than for any of the other species and anthropogenic variables dominated the models. Silktree was most likely to occur close to roads (<250 m) but in areas of lower human population (<2250 people km^{-2}), and at lower elevations, but not along the coast (50 to 550 m), with moderate precipitation (1100 to 1550 mm), moderate minimum temperature of coldest month (-5°C to 4°C) and low precipitation seasonality (10 to 30% variation) (Table 6-3). As the silktree model was driven by anthropogenic variables the application of climate change scenarios did not have such a strong influence, with only a slight increase predicted under scenarios B and C (Table 6-4). Silktree had 53 percent of forest predicted with no occurrence and 21 percent of forest were predicted to be invaded (Figure 6-6, Table 6-5). This model had the highest disagreement between models, with 26% of the forest predicting occurrence with two or three models.

6.4.4 Princesstree

Princesstree had an acceptable model with a test AUC of 0.85 and test omission rate of 0.25 (Table 6-2). The model identified the northern part of the region as having high invasion probability (Figure 6-4). There are a couple of low density points that are missed in the southern part of the study area, noticeable when comparing the FIA county map (Figure 6-4) with potential distribution maps. Minimum temperature of coldest month was the most dominant variable, contributing 54 percent to the model. Princesstree was most likely to occur at moderate to low minimum temperature in coldest month (-7°C to -1°C), at low precipitation seasonality (<18% variation), where there is also moderate precipitation in the warmest quarter (270 to 370 mm), with lower human population (<300 people km^{-2}) and more likely to less pine forest (<20%) (Table 6-3).Under all climate change scenarios the potential of forest will be reduced from the current potential but still far exceeds the proportion of the forest currently invaded (Table 6-4). Princesstree had the most definitive models when combined with 77 percent of the forest predicted not to have princesstree, and only 15 percent predicted to have princesstree (Figure 6-6, Table 6-5).

6.4.5 Tallowtree

Tallowtree had the strongest model of all species with test AUC of 0.92 and test omission rate of 0.12 (Table 6-2). It also had the highest current occurrence (3%) and lowest potential occurrence (12%). The model identified the northern part of the region as having high invasion probability (Figure 6-5). Minimum temperature of coldest month was the most dominant variable, contributing 63 percent to the model. Tallowtree was most likely to occur at moderate to high minimum temperature in coldest month (>1°C), at moderate precipitation seasonality (10 to 55), and more likely to have less forest (<65%) (Table 6-3). Under all scenarios apart from scenario D, the distribution of tallow tree increases dramatically (Table 6-4). Sixty four percent of the area was predicted to have no tallow tree, with occurrence predicted in 11 percent of the forests (Figure 6-6, Table 6-5).

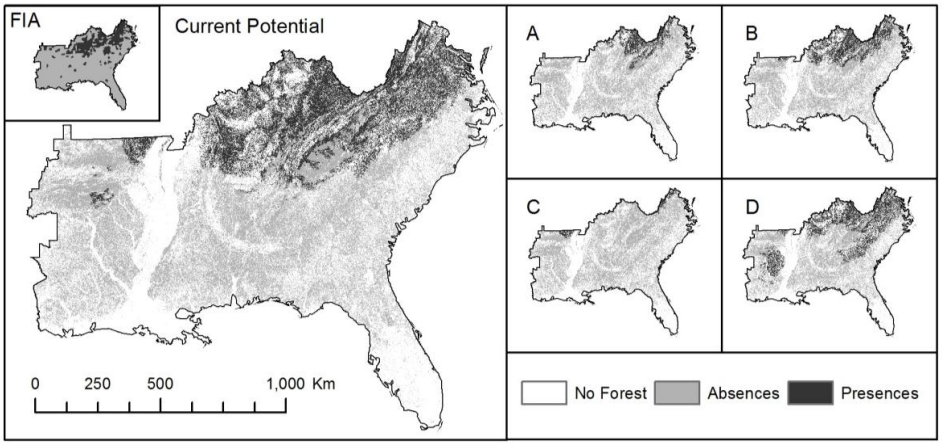

Figure 6-1: Potential invasion probability maps for tree of heaven in the Southeastern United States under five climate scenarios.

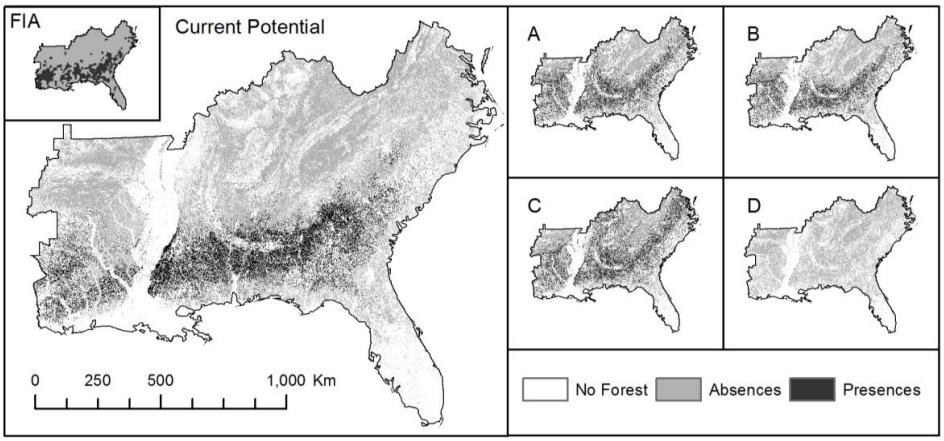

Figure 6-2: Potential invasion probability maps for chinaberry in the Southeastern United States under five climate scenarios.

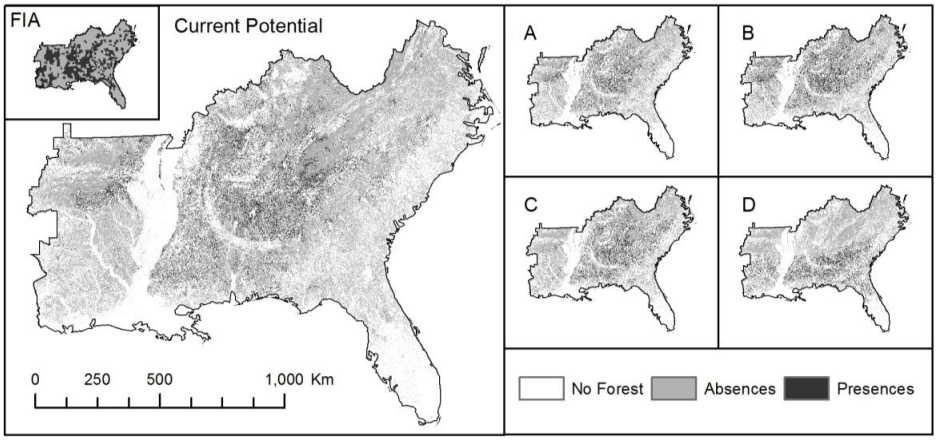

Figure 6-3: Potential invasion probability maps for silktree in the Southeastern United States under five climate scenarios.

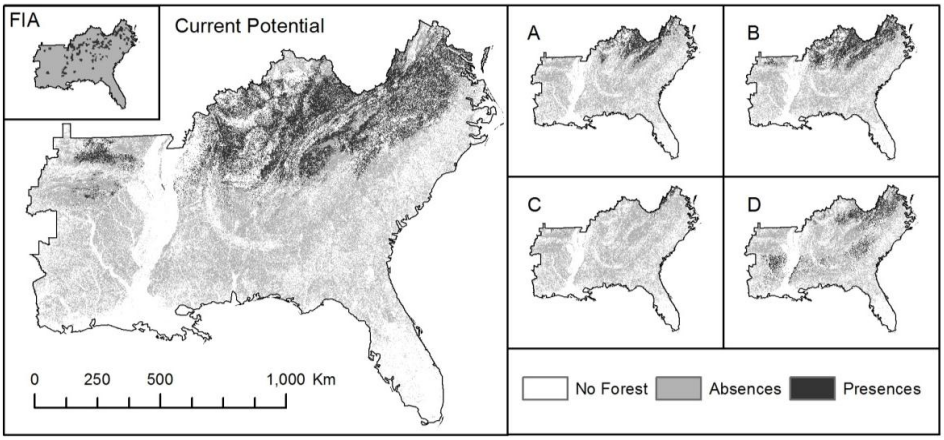

Figure 6-4: Potential invasion probability maps for princesstree in the Southeastern United States under five climate scenarios.

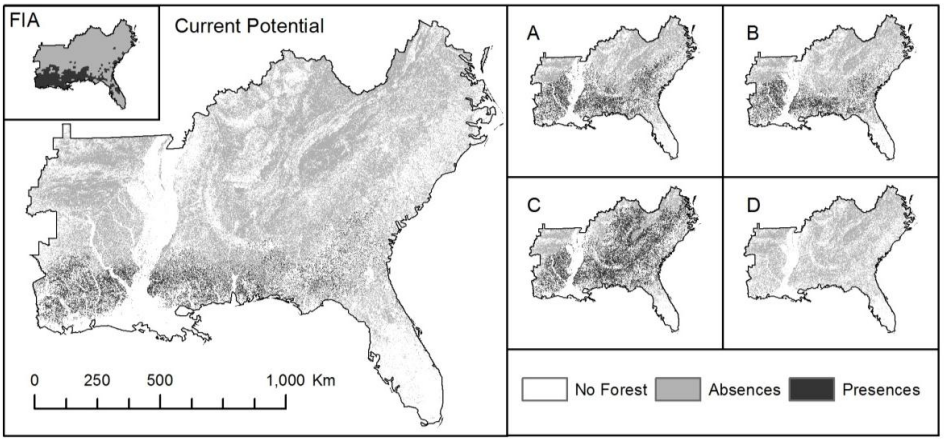

Figure 6-5: Potential invasion probability maps for tallowtree in the Southeastern United States under five climate scenarios.

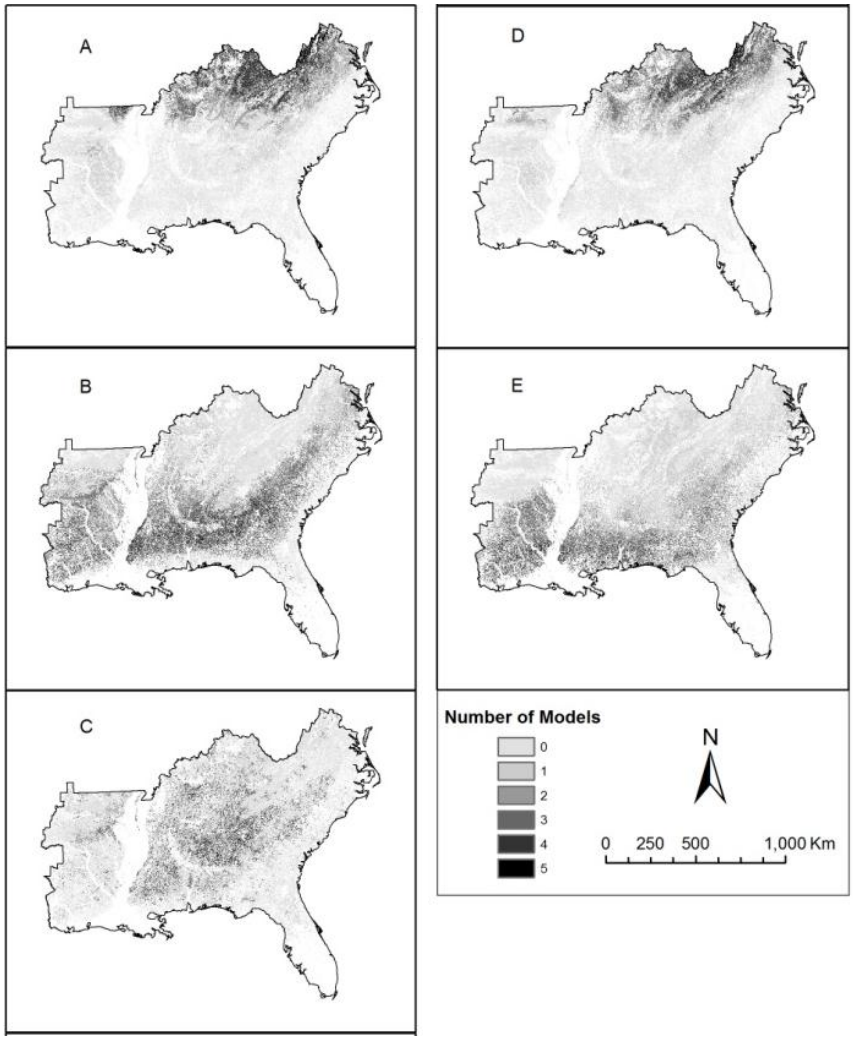

Figure 6-6: Potential invasion probability maps for five invasive trees in the Southeastern United States under five climate scenarios (A – Tree of Heaven, B - Chinaberry, C – Silktree, D – Princesstree, E – Tallowtree).

6.5 Discussion

Of the five invasive tree species modelled, all are predicted to significantly increase their distribution at the regional level under current conditions, and also increase under all climate change scenarios, in some cases dramatically, and in others minimally. This agrees with other studies that suggest that the distribution of alien plants will increase This may have consequences to long term structure of forests in the region, and greatly impact potential management options (Martin et al. 2009; Miller et al. 2012).Increase under many of the potential future conditions (Elith et al. 2010; Hellmann et al. 2008; Walther et al. 2009). Other studies have used climate based distribution models to predict risk of invasion (Herborg et al. 2007;

Nuñez & Medley 2011; Zambrano & Martínez-Meyer 2006). However, there are a number of constraints in the modelling approach that we have taken. The main one was the use of environmental information for modelling species occurrence from a realized niche, which may lead to underestimating its potential distribution. This is primarily due to biotic interactions and dispersal limitations that do not allow the species to inhabit some environmental conditions that might otherwise be suitable. This is of particular concern with invasive species that are unlikely to be at equilibrium with their environment. This problem would be more pronounced in models with a lower contribution for landscape level variables such as climate. The more a model is driven by anthropogenic variables, such as roads, the more likely the species is in the early stages of invasion process, as these anthropogenic variables are likely to influence the spread. This can be seen in the silktree model in which 57 percent of the explained variation of the model was due to of anthropogenic variables. In some cases, it is the disturbance that results from the anthropogenic activities that creates habitat suitable for the species to flourish. With the increasing population in the southeastern United States, this disturbance will likely increase, thus increasing the distribution of alien species.

Models dominated by environmental variables will be more stable over time under the same climatic conditions. Such models will also be more useful to assessing differing climate change scenarios. One very strong model that was developed was for tallowtree. There are two main reasons for the high explanatory power of this model: first, tallowtree occurred in three percent of the plots, the most of the invasives examined; and second, the model was highly dominated by environmental variables, suggesting tallowtree has been established long enough to be spreading throughout the range and not limited by human activity. Tallowtree has recently been modelled in a number of studies (Gan et al. 2009). Gan and others (2009) focused on Texas and Mississippi, where current invasion is the greatest. Our models predicted that 12 percent of the region's forests have the potential to be invaded under current conditions. As a subtropical-to-temperate species, tallowtree is likely to be limited in its northern range by minimum temperatures (Dirr 1998). Gan and others (2009) also found elevation and minimum temperature to be prime variables. This is concerning as climate change is expected to raise temperatures, allowing invasion further northward. With three of the four climate change scenarios applied the potential for invasion dramatically increases, with up to 64 percent of forest becoming potentially invasible under scenario C.

Tree of heaven also had a model strongly dominated by climatic conditions, specifically mean minimum temperature and rainfall seasonality, suggesting that it may already occur throughout its current potential range, though still increasing in prevalence. In the southeastern region tree of heaven's current distribution is limited by temperature, with a preference for areas with cooler temperatures. Under future conditions it may no longer be a species of concern in the region. Princesstree and chinaberry also had models where climatic variables dominated, however with fewer data points, higher omission rates and lower AUC's the ranges may be limited and predictions weakened, by to inadequate definition of the potential distributions (Menke et al. 2009).

The five species responded differently to the future scenarios with species models highly dependent on climate variables differing more among current conditions and future scenarios. The variation among climate change scenarios for one species is greater than the variation between species under current conditions. This suggests the importance of adaptive management and long term monitoring programs and the need for further development in determining future conditions. To further increase the application of SDM integrating potential scenarios with the likelihood estimates would be beneficial. A recent paper by Elith and others (2010) suggested that it is essential to think about the biology of the species, integrate the fitted models and reduce as much as possible unwanted effects. This is the approach taken in this study within the limitations imposed by our knowledge of these species and data constraints. However, further study into species response with environmental constraints would be beneficial. SDM can be a useful tool in assessing potential impacts of invasive species and the impact that future conditions may have, but is currently limited by our knowledge of the species responses and uncertainty of future conditions. SDMs are however a useful start for informing our knowledge and, with some refinement, will become a key management tool.

6.6 Literature cited

Anderson, J.R., Hardy, E.E., Roach, J.T. & Witmer, R.E. (1976) A land use and land cover classification system for use with remote sensor data. United States Geological Survey Professional Paper 964, United States Government Printing Office, Washington, D.C.

Anderson, R.P. & Raza, A. (2010) The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with

montane rodents (genus *Nephelomys*) in Venezuela. Journal of Biogeography, 37, 1378-1393.

Avon, C., Bergès, L., Dumas, Y. & Dupouey, J.L. (2010) Does the effect of forest roads extend a few meters or more into the adjacent forest? A study on understory plant diversity in managed oak stands. Forest Ecology and Management, 259, 1546-1555.

Bechtold, W.A. & Patterson, P.L. (2005) The enhanced forest inventory and analysis program — national sampling design and estimation procedures. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station.

Birdsey, R.A. & Schreuder, H.T. (1992) An overview of forest inventory and analysis estimation procedures in the eastern United States – with an emphasis on the components of change. USDA Forest Service General Technical Report RM-214.

Bruce, K.A., Cameron, G.N., Harcombe, P.A. & Jubinsky, G. (1997) Introduction, impact on native habitats, and management of a woody invader, the Chinese tallow-tree , *Sapium sebiferum (*L.) Roxb. Natural Areas Journal, 17, 255–260.

Carter, W.K. (2007) Tree seedling and sapling density and deer browsing incidence on recently logged and mature non-industrial private forestlands in Virginia, USA. Forest Ecology and Management, 242, 671–677.

Chapman, E.L., Chambers, J.Q., Ribbeck, K.F., Baker, D.B., Tobler, M.A., Zeng, H. & White, D.A. (2008) Hurricane Katrina impacts on forest trees of Louisiana's Pearl River Basin. Forest Ecology and Management, 256, 883-889.

Conway, W.C., Smith, L.M. & Bergan, J.F. (2002) Potential allelopathic interference by the exotic Chinese tallow tree (*Sapium sebiferum*). American Midland Naturalist, 148, 43–53.

Creager, R.A. (1992) Seed germination, physical and chemical control of catclaw mimosa (*Mimosa pigra* var. *pigra*). Weed Technology, 6, 884-891.

Dirr, M.A. (1998) Manual of woody landscape plants: their identification, ornamental characteristics, culture, propagation and uses. Stipes Publishing, Champaign, IL.

Edwards Jr., T., Cutler, D., Zimmermann, N., Geiser, L. & Moisen, G. (2006) Effects of sample survey design on the accuracy of classification tree models in species distribution models. Ecological Modelling, 199, 132-141.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. Overton, J., Townsend Peterson, A.,

Phillips, S.J., Richardson, K., Scachetti-pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography, 29, 129-151.

Elith, J., Kearney, M. & Phillips, S. (2010) The art of modelling range-shifting species. Methods in Ecology and Evolution, 1, 330-342.

Elton, C. (1927) Animal Ecology. Chicago, University of Chicago Press.

Evans, M.C., Watson, J.E.M., Fuller, R. A., Venter, O., Bennett, S.C., Marsack, P.R. & Possingham, H.P. (2011) The spatial distribution of threats to species in australia. BioScience, 61, 281-289.

Gan, J., Miller, J.H., Wang, H. & Taylor, J.W. (2009) Invasion of tallow tree into southern US forests: influencing factors and implications for mitigation. Canadian Journal of Forest Research, 39, 1346-1356.

Gelbard, J.L. & Belnap, J. (2003) Roads as conduits for exotic plant invasions in a semiarid landscape. Conservation Biology, 17, 420-432.

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M. & Tyler, D. (2002) The national elevation dataset. Photogrammetric Engineering and Remote Sensing, 68, 5-11.

Graham, C.H. & Hijmans, R.J. (2006) A comparison of methods for mapping species ranges and species richness. Global Ecology and Biogeography, 15, 578-587.

Griep, M.T. & Collins, B. (2012) Southern Forest Futures Project Chapter 14: Wildlife and Forest Communities. In Press.

Heisey, R.M. (1996) Identification of an allelopathic compound from *Ailanthus altissima* (Simaroubaceae) and characterization of its herbicidal activity. American Journal of Botany, 83, 192.

Hellmann, J.J., Byers, J.E., Bierwagen, B.G. & Dukes, J.S. (2008) Five potential consequences of climate change for invasive species. Conservation Biology, 22, 534-543.

Herborg, L., Jerde, C. & Lodge, D. (2007) Predicting invasion risk using measures of introduction effort and environmental niche models. Ecological, 17, 663-674.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25, 1965-1978.

Hosmer, D.W. & Lemeshow, S. (2000) Applied Logistic Regression. Wiley-Interscience, New York.

Huebner, C. (2003) Vulnerability of oak-dominated forests in West Virginia to invasive exotic plants: temporal and spatial patterns of nine exotic species using herbarium records and land classification data. Castanea, 68, 1-14.

Hunt, K.W. (1947) The Charleston woody flora. American Midland Naturalist, 37, 670-756.

Hutchinson, M.F. (1987) Methods for generation of weather sequences. In Bunting, A.H. (Ed.) Agricultural environments: characterisation, classification and mapping. Wallingford, CAB, International, 149-157.

Iverson, L., Prasad, A. & Matthews, S. (2007) Modeling potential climate change impacts on the trees of the northeastern United States. Mitigation and Adaptation Strategies for Global Change, 13, 487-516.

Iverson, L., Prasad, A. & Schwartz, M.W. (1999) Modeling potential future individual tree-species distributions in the eastern United States under a climate change scenario: a case study with *Pinus virginiana*. Ecological Modelling, 115, 77–93.

Jeschke, J.M. & Strayer, D.L. (2008) Usefulness of bioclimatic models for studying climate change and invasive species. Annals of the New York Academy of Sciences, 1134, 1-24.

Jiménez-Valverde, A. (2011) Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. Global Ecology and Biogeography, 21, 498-507.

Jiménez-Valverde, A. & Lobo, J.M. (2006) The ghost of unbalanced species distribution data in geographical model predictions. Diversity and Distributions, 12, 521-524.

Johnson, J.E., Mitchem, D.O. & Kreh, R.E. (2003) Establishing royal paulownia on the Virginia Piedmont. New Forests, 25, 11–23.

Knapp, L.B. & Canham, C.D. (2000) Invasion of an old-growth forest in New York by *Ailanthus altissima*: Sapling growth and recruitment in canopy gaps. Journal of the Torrey Botanical Society, 127, 307-315.

Kowarik, I. & Saumel, I. (2007) Biological flora of Central Europe: *Ailanthus altissima* (Mill.) Swingle. Perspectives in Plant Ecology, Evolution and Systematics, 8, 207-237.

Langdon, K.R. & Johnson, K.D. (1994) Additional notes on invasiveness of *Paulownia tomentosa* in natural areas. Natural Areas Journal, 14, 139-140.

Lawrence, J.G., Colwell, A., Sexton, O.J., Journal, A. & Jul, N. (1991) The Ecological Impact of Allelopathy in *Ailanthus altissima* (Simaroubaceae). American Journal of Botany, 78, 948-958.

Lemke, D., Hulme, P.E., Brown, J.A. & Tadesse, W. (2011) Distribution modelling of Japanese honeysuckle (*Lonicera japonica*) invasion in the Cumberland Plateau and Mountain Region, USA. Forest Ecology and Management, 262, 139-149.

Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R. (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography, 28, 385-393.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography, 17, 145-151.

Longbrake, A.C.W. (2001) Ecology and invasive potential of *Paulownia tomentosa* (Scrophulariaceae) in hardwood forest landscape. PhD, Ohio University, Plant Biology, 174.

Longbrake, A.C. & McCarthy, B.C. (2001) Biomass allocation and resprouting ability of princess tree (*Paulownia tomentosa*: Scrophulariaceae) across a light gradient. The American Midland Naturalist, 146, 388-403.

Mack, M.C. & D'Antonio, C.M. (1998) Impacts of biological invasions on disturbance regimes. Trends in Ecology & Evolution, 13, 195-198.

Manel, S., Williams, H.C. & Ormerod, S.J. (2002) Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology, 38, 921-931.

Martin, M.R., Tipping, P.W. & Sickman, J.O. (2009) Invasion by an exotic tree alters above and belowground ecosystem components. Biological Invasions, 11, 1883-1894.

Menke, S.B., Holway, D.A., Fisher, R.N. & Jetz, W. (2009) Characterizing and predicting species distributions across environments and scales: Argentine ant occurrences in the eye of the beholder. Global Ecology and Biogeography, 18, 50-63.

Miller, J.H. (1990) *Ailanthus altissima* (Mill.) Swingle. (eds R.M. Burns & B.H. Honkala), p. 877. U.S. Department of Agriculture, Forest Service, Washington, DC.

Miller, J.H., Chambliss, E.B. & Loewenstein, N.J. (2010) A field guide for the identification of invasive plants in southern forests. Gen. Tech. Rep. SRS–119. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 126 p.

Miller, J.H., Lemke, D. & Coulston, J. (2012) The southern forest futures project, Chapter 15. The invasion of southern forests by nonnative plants: current and future occupation with impacts, management strategies, and mitigation approaches. In Press.

Nuñez, M. A. & Medley, K. A. (2011) Pine invasions: climate predicts invasion success; something else predicts failure. Diversity and Distributions, 17, 703-713.

PRISM (2007) http://www.prismclimate.org. PRISM Group, Oregon State University.

Pattison, R.R. & Mack, R.N. (2009) Environmental constraints on the invasion of *Triadica sebifera* in the eastern United States: an experimental field assessment. Oecologia, 158, 591-602.

Pennington, T.D. (1981) A Monograph of Neotropical Meliaceae. Flora Neotropica, 28, 1-470.

Peterson, A.T. (2003) Predicting the geography of species invasions via ecological niche modeling. Quarterly Review Biology, 78, 419-433.

Peterson, A.T. & Vieglais, D.A. (2001) Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. BioScience, 51, 363.

Phillips, S., Anderson, R. & Schapire, R. (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190, 231-259.

Pimentel, D., Zuniga, R. & Morrison, D. (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecological Economics, 52, 273-288.

Pysek, P. (2004) The most complete global overview of invasive species in natural areas. Diversity and Distributions, 10, 505-506.

Rejmanek, M., Pitcairn, M.J. & Rejmánek, M. (2002) When is eradication of exotic pest plants a realistic goal? Turning the tide: the eradication of invasive species: proceedings of the International Conference on Eradication of Island Invasives. (eds C.R. Veitch & M.N. Clout), pp. 249–253. IUCN SSC Invasive Species Specialist Group., Cambridge, UK.

Renne, I.J., Barrow, W.C., Johnson Randall, L.A. & Bridges, W.C. (2002) Generalized avian dispersal syndrome contributes to Chinese tallow tree (*Sapium sebiferum*, Euphorbiaceae) invasiveness. Diversity and Distributions, 8, 285-295.

Ricciardi, A., Palmer, M.E. & Yan, N.D. (2011) Should biological invasions be managed as natural disasters? BioScience, 61, 312-317.

Rogers, W.E. & Siemann, E. (2005) Herbivory tolerance and compensatory differences in native and invasive ecotypes of chinese tallow tree (*Sapium sebiferum*). Plant Ecology, 181, 57-68.

Schwartz, M.W., Iverson, L.R. & Prasad, A.M. (2001) Predicting the potential future distribution of four tree species in Ohio using current habitat availability and climatic forcing. Ecosystems, 4, 568-581.

Sharma, G.P., Singh, J.S. & Raghubanshi, A.S. (2005) Plant invasions: emerging trends and future implications. Current, 88, 726-734.

Smith, W.B., Miles, P.D., Perry, C.H. & Pugh, S.A. (2009) Forest resources of the United States 2007. Gen. Tech. Rep. WO-78. Washington, DC.

Stohlgren, T.J., Chong, G.W., Schell, L.D., Rimar, K. A., Otsuki, Y., Lee, M., Kalkhan, M. A. & Villa, C. A. (2002) Assessing vulnerability to invasion by nonnative plant species at multiple spatial scales. Environmental Management, 29, 566-577.

Thuiller, W. (2007) Biodiversity: Climate change and the ecologist. Nature, 448, 550-552.

Tourn, G., Menvielle, M., Scopel, A. & Pidal, B. (1999) Clonal strategies of a woody weed: *Melia azedarach*. Plant and Soil, 217, 111-117.

USBOC (2000) United States Bureau of the Census – Tiger Files, Geography Division.

USDA (2011) PLANTS National Plants Database. National Plant Data Center, Baton Rouge. http://plants.usda.gov.

USDA Forest Service. (2008) Nonnative invasive species of southern forests and grasslands, compiled by a southern regional task force for the assessment of nonnative invasive species. http://www.invasive.org/south/taskforce.html.

Vandermeer, J. (1972) Niche theory. Annual Review of Ecology and Systematics, 3, 107-132.

Walther, G.-R., Roques, A., Hulme, P.E., Sykes, M.T., Pysek, P., Kühn, I., Zobel, M., Bacher, S., Botta-Dukát, Z., Bugmann, H., Czúcz, B., Dauber, J., Hickler, T., Jarosík, V., Kenis, M., Klotz, S., Minchin, D., Moora, M., Nentwig, W., Ott, J., Panov, V.E., Reineking, B., Robinet, C., Semenchenko, V., Solarz, W., Thuiller, W., Vilà, M., Vohland, K. & Settele, J. (2009) Alien species in a warmer world: risks and opportunities. Trends in Ecology & Evolution, 24, 686-693.

Wang, H.-H., Grant, W.E., Swannack, T.M., Gan, J., Rogers, W.E., Koralewski, Tomasz E. Miller, J.H. & Taylor Jr, J.W. (2011) Predicted range expansion of Chinese tallow tree (Triadica sebifera) in forestlands of the southern United States. Diversity and Distributions, 17, 552-565.

Wear, D. & Greis, J. (2012) Summary report the southern forest futures project. In Press.

Wear, D.N. & Greis, J.G. (2002) Southern forest resource assessment: summary of findings. Journal of Forestry, 100, 6–14.

Webb, C.J., Sykes, W.R. & Garnock-Jones, P.J. (1988) Flora of New Zealand. Botany Division, DSIR, Christchurch.

Weber, E. (2003) Invasive plants of the World. CABI Publishing, CAB International, Wallingford, UK.

Wilcove, D.S., Rothstein, D., Dubow, J., Phillips, A. & Losos, E. (2000) Leading threats to biodiversity: what's imperilling US species. Precious Heritage The Status of Biodiversity in the United States Oxford University Press Oxford, 239–254.

Williams, C.E. (1993) The exotic empress tree, *Paulownia tomentosa*: an invasive pest of forests? Natural Areas Journal, 13, 221-222.

Wilson, J.R.U., Richardson, D.M., Rouget, M., Procheş, S., Amis, M.A., Henderson, L. & Thuiller, W. (2007) Residence time and potential range: crucial considerations in modelling plant invasions. Diversity and Distributions, 13, 11–22.

Zambrano, L. & Martínez-Meyer, E. (2006) Invasive potential of common carp (*Cyprinus carpio*) and Nile tilapia (*Oreochromis niloticus*) in American freshwater systems. Canadian Journal of Fisheries and Aquatic Sciences, 1910, 1903-1910.

Zimmermann, N.E., Edwards, T.C., Moisen, G.G., Frescino, T.S. & Blackard, J. a. (2007) Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. The Journal of Applied Ecology, 44, 1057-1067.

CHAPTER 7.

Data for Invasive Plant Distribution Modelling: importance of scale, collection and selection.

Collaborators: *John Coulston and James Miller*

7.1 Abstract

Invasive species are now recognized as a major component of global environmental change. They have a tremendous cost to both the economy and environment; in the United States invasive species cause losses in excess of US $120 billion a year. Invasive species are one of the fundamental ecological challenges we are currently facing. One tool that can assist in assessing and managing is species distribution models (SDMs). SDMs make predictions about unsampled locations using the relationships between species and environmental parameters. They are designed to make use of known information of species that are at equilibrium with the environment, and work best when a full range of the environment has been sampled. However, they have also shown promise for the study of invasive species, species that have yet to reach equilibrium, or are not fully sampled. In this chapter we assess the applicability of different types of datasets to modelling five invasive trees in forests of the southeastern United States. We used intensively collected systematically acquired absence/presence data (Forest Inventory and Analysis, FIA), occurrence only, opportunistically acquired data (Global Biodiversity Information Facility, GBIF), and combined datasets. The resulting potential distribution of the five invasive tree species in the southeastern region depended on the source and scale (global and regional) data used to make the predictions. Overall, we showed large variation in modelled distribution, with models built on global opportunistically acquired data always having larger geographical ranges. For each species the potential distribution using global GBIF data presented less error as assessed by AUC and omission rate, but suggested between two to four times greater areas of potential distribution (51 to 99%) when compared with models based on the regional FIA data (12 – 44%) and regional GBIF data (23 – 41%). Given the limitations of only using climate and elevation data we would suggest that the global GBIF models represent a

broader range of conditions that may not be limited by the same biotic and anthropogenic constraints as in the southeastern region.

Comparison of regional GBIF and FIA models showed FIA to have stronger models, with low spatial congruence between the models for some of the species. In general, there were more FIA points than GBIF regional points but this was not true for all species, thus the number of data points cannot be the only component driving the more accurate FIA models. The FIA data points were measured within metres of their true location, the GBIF data is at an unknown level of accuracy; however, in some cases the GBIF points seem to be based on county centroids (potential 100's of kilometres from where the species was found), this could be another component of the greater accuracy found in the FIA models. A number of studies have used an ensemble approach, using a number of modelling techniques. In this case we took an ensemble approach with models built with different datasets. The strongest regional models were in general the ones derived from the combined FIA and regional GBIF data, suggesting that though the quality of data is important, quantity is fundamental.

Temperature in the coldest month was the only variable used in all models, with an average contribution of 35 percent. Annual precipitation and precipitation seasonality were the only variables that contributed, on average, more than ten percent to the models. The overall fit of the models suggests that species distribution for these species is climate driven but at a regional level there are maybe other variables influencing the distribution.

SDMs can be used to assist in identifying areas of greatest risk of invasion and in developing management strategies, but they are not the only tool and need to be used in context of the data that was used in their development.

7.2 Introduction

Invasive species pose a major threat to the sustainability of natural ecosystems. They are one of the greatest threats to biodiversity, second only to habitat destruction (Wilcove et al. 1998), and a major component of global environmental change (Ricciardi 2007; Vitousek et al. 1997). In the United States, invasive species cause environmental damage and economic loss in excess of US$120 billion a year, $34 billion per year for plants alone (Pimentel et al. 2005). Controlling invasive plants is costly and affects both the viability of the forest industry as well as ecosystem sustainability (Pimentel et al. 2005). There are over 380 recognized invasive plants in southeastern United States forests and grasslands, with 53 ranked as high-to-medium risk to

natural communities (Miller et al. 2010). Identifying areas of potential invasion is an important part of natural ecosystem management, and species distribution models (SDM) can help. SDMs combine concepts from ecology and natural history with more recent developments in statistics and geospatial information systems (Elith & Leathwick 2009). They are the core of predictive geographical modelling in ecology (Guisan & Zimmermann 2000). SDMs rely on the niche concept that emphasizes species requirements, represented by predominately abiotic factors that control species distribution (Franklin 2009). Ideally SDMs are developed using comprehensive well-designed surveys with absence and presence data from the full geographic range and for locations where the species are near equilibrium with their environment (Franklin 2009). SDMs have been used for biodiversity assessment, habitat management, conservation planning, viability analysis, ecosystem modelling and assessment of the potential impact of climate change. One area in which SDMs may have great value is the assessment of the potential distribution of invasive species (Peterson & Vieglais 2001; Peterson 2003; Andersen et al. 2004). However, the explanatory power of models can be limited when species are not at equilibrium with their environment; models are extrapolated in time or space beyond the limit of the data, or when inadequate data is used (Elith & Leathwick 2009), all of which are potential issues with invasive SDMs (Phillips et al. 2009; Hulme & Weser 2011). Thus data sourcing, availability and quality are fundamental components to the development of useful and accurate SDMs.

Datasets can be grouped in two basic types; opportunistically acquired, which were multiple source data with low quality assurance and quality control (QA/QC), and systematically acquired which were intensively surveyed data with high QA/QC. Opportunistically acquired data is generally cheaper and with a greater coverage area whereas systematically acquired data is generally more time consuming and costly to acquire, resulting in application at small scales. The applicability of these different datasets has only recently become a subject of discussion. In the last few years the availability and accessibility of species occurrence data has dramatically increased. Along with ease of access to data, user friendly graphic interface modelling packages have led to a proliferation of SDM use, as well as potential misuse.

There is an immense amount of opportunistically acquired data on the occurrence of species in museums, private collections, herbaria and in literature now available through several internet portals making this data freely available. One of these portals is the Global Biodiversity Information Facility (GBIF), which contains species locations throughout the globe. The mission of GBIF is to facilitate free and open access to worldwide biodiversity data, via the internet to

support sustainable development (www.gbif.org). GBIF data is not collected systematically and relies on information uploaded from partner organisations (Flemons et al. 2007). As a consequence, GBIF makes no claim regarding the accuracy or reliability of the data or for its suitability to particular applications. Nevertheless, these global data are increasingly being used to model regional species distributions (Gallagher et al. 2010; Kruijer et al. 2010; Bystriakova & Schneider 2011; Nori et al. 2011). Opportunistically acquired data shows three major types of bias: spatial, taxonomic and temporal (Soberón et al. 2000). Opportunistically acquired data can also contain extensive errors in locations and species identification (Graham et al. 2004). Thus, data needs to be used with caution, and with an understanding of its limitations. The usefulness of records for distribution modelling predominantly depends on the precision of the location, thus it is necessary to check very thoroughly for georeferencing errors in opportunistically acquired data, by checking for obviously outlying points (Williams et al. 2002). SDMs are designed to extrapolate from incomplete data, but they assume that the species data is representative of the environments occupied. This is an issue within some geographical areas with sampling often less intensive in the tropics and arid environments, and much more intensive in temperate areas. Even within regions and countries, sampling is often spatially uneven with locations occurring closer to roads, rivers and cities (Hijmans et al. 2000; Soberón et al. 2000; Reddy & Dávalos 2003; Sánchez-Fernández et al. 2008). One further limitation of opportunistically acquired data is the lack of records of species absence to complement records of species presence (Graham et al. 2004).

In contrast to this, systematically acquired data normally contains both absence and presence information; it is more accurately located, and is normally designed around a statically sound sampling strategy. However, this approach is generally cost prohibitive at larger scales unless it is part of a broader national sampling program. Thus, most SDMs developed at a regional or global scale utilise opportunistically acquired data. In the United States the Forest Service has a national data collection program that is used to assess the nation's forests, the Forest Inventory and Analysis program (FIA). The FIA Program aims are to collect, analyse and report information on the status, trends and condition of America's forests based on a statically sound sampling design. It is a single inventory program that includes all forested lands in the United States, regardless of ownership or availability for forest harvesting. In the southeastern region there are over 50,000 FIA plots, with recent forest inventories being completed every five to seven years, and approximately 20 percent of these points assessed every year (USDA/FS 2007).

It has very little spatial, taxonomic or temporal bias, and contains both presence and absence data making it extremely useful for regional scale SDMs. FIA data has been widely applied to SDMs, developing strong regional models (Edwards et al. 2006; Zimmermann et al. 2007; Lemke et al. 2011; Wang et al. 2011).

Beyond data source, there are a number of other issues that arise in data set selection for SDMs, one of which is the scale at which data is sourced. This area of modelling has been little explored, in terms of the application and relevance of global scale datasets to regional scale models. Recently, Sanchez-Fernandez and others (2011) examined this issue and suggested that models derived solely from regional data sets may underestimate the potential distribution. Other studies have found little difference in geographically biased models (Wolmarans et al 2010). This then opens the door to the discussion of whether the drivers at regional scale and global scale are the same. There is no single scale at which ecological patterns should be studied (Levin 1992), however, the patterns may be influenced by differing characteristics at different scales. Climate tends to dominate distributions at the global scale, whereas at local and eco-regional scale (a few to hundreds of kilometres) topography, geology and land use give variation in climate, nutrient and moisture availability that influence species distribution (Mackey & Lindenmayer 2001).

In this chapter we focused on three questions specific to the application of SDMs; 1) How does scale and data collection influence regionally modelled potential distribution? 2) Do environmental determinants remain consistent across scale and datasets? 3) Can value be added by integrating SMDs from different datasets? To address these questions we developed SMDs for five trees invasive to the forests of the southeastern region of the United States using regional systematically acquired data, regional opportunistically acquired data and global opportunistically acquired data. These five tree species represent varying stages of invasion, all species are unlikely to have reached equilibrium (Miller et al. 2012). We assessed the variation in models based on species datasets using both global data, including native ranges and more intensive, systematically acquired data collection limited to just the southeastern region.

7.3 Methods

7.3.1 Study Region

This study focuses on the distribution of five invasive trees in the forested areas of the southeastern region of the United States. The southeastern region is defined as 13 states (Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, east Oklahoma, South Carolina, Tennessee, east Texas, and Virginia) and crosses temperate, sub-tropical, tropical, and arid climate zones (Hijmans et al. 2005). Elevation for this region ranges from sea level to 1960 m and rainfall ranges from 570 mm to 2070 mm (Hijmans et al. 2005). The mean annual temperatures are between 6 and 24 ^{o}C, with the average maximum summer temperatures ranging from 20 to 36 ^{o}C, and average minimum winter temperatures between -10 and 16 ^{o}C (Hijmans et al. 2005). The landscape is heavily influenced by human use with the majority of forest having been harvested at least once in the last two centuries (Wear & Greis 2002). Forest remains the dominant land cover of this region at forty percent (Smith et al. 2009), with the most common forest communities being oak/hickory, loblolly/shortleaf pine, oak/pine, and oak/gum/cypress (Wear & Greis 2002). Forests develop more rapidly in the southeastern region than in other regions in the United States, partly because of humid temperate and subtropical climates and partly because of the relatively fast growth rates of native tree species. The southeastern region produces nearly six percent of the timber production in the United States (Wear & Greis 2012). Along with the important economic value of the southeastern forests, they are among the most biologically diverse temperate forests in the world (Wear & Greis 2002).

7.3.2 Study Species

We focus on the five most widespread invasive trees in the forests of the southeastern region: tree of heaven (*Ailanthus altissima,* [Mill.] Swingle), chinaberry tree (*Melia azedarach*, L.), silktree (*Albizia julibrissin,* Durazz.), princesstree (*Paulownia tomentosa,* [Thunb.] Siebold & Zucc. Ex Steud.) and tallowtree (*Triadica sebifera,* [L.] Small). All five species were introduced to the region from Asia in the late 18th century or early 19th century as ornamentals or for commercial use (Miller et al. 2010; Miller et al. 2012) and are now found through the majority of the eastern and southeastern United States (PLANTS 2011). They share several traits that have facilitated their widespread distribution (Table 7-1), including early reproduction, multiple reproductive modes and human use.

Table 7-1: Summary of invasive species (Flemons et al. 2007; Miller et al. 2010, 2011; PLANTS 2011; USDA/FS 2007)

Scientific name	Ailanthus altissima	Melia azedarach	Albizia julibrissin,	Paulownia tomentosa	Triadica sebifera
Common name	Tree of Heaven	Chinaberry	Silktree	Princesstree	Tallowtree
Area of origin	China	China/India	Asia	China	Asia
Current occurrence in FIA plots	2.5%	1.4%	2.0%	0.7%	3.2%
Number of US states with occurrence	43	19	29	27	10
Latitudinal range based on all GBIF data	55^o to 18.5^o & -18.5^o to -55^o	47^o to -47^o	50.5^o to 21.5^o & -21.5^o to -50.5^o	53^o to 26.5^o & -26.5^o to -53^o	40.5^o to 20.5^o & -20.5^o to -40.5^o
Estimated date of entry into United States	1784	1830	1785	1844	1900
Reproductive mode	vegetatively and seed	vegetatively and seed	vegetatively and seed	vegetatively and seed	vegetatively and seed
Age of reproduction	2-4 yrs	4-6 yrs	3-5 yrs	5-7 yrs	3-5 yrs
Life span	30-50 yrs	20-30 yrs	30-45 yrs	70-125 yrs	50-80 yrs
Shade tolerance	moderate	moderate	moderate	high	low
Cold tolerance	high	low	moderate	moderate	low
Human use	ornamental	ornamental	ornamental	ornamental / timber	ornamental / industry

7.3.3 Plant occurrence data

To examine the influence of data quantity, quality and geographical relevance, data at both the global (Global Biodiversity Information Facility, GBIF) and regional scale (USDA Forest Inventory and Analysis, FIA) were used to generate predicted potential distributions in the forested areas of the southeastern region (Figure 7-1).

For opportunistically acquired data we downloaded all available data from GBIF for each of the five species (including both native and introduced ranges), and then removed all repeat records, data without location information and points that were recorded with less than a tenth of degree accuracy (approximately 10 km). Information was rarely given as to whether a species was naturally occurring, invasive or planted, only that it occurred, thus all data that had an location recorded with less than a tenth of degree accuracy were retained for further analysis (GBIF gives no assessment of location accuracy). Location information was not limited to forested land use.

An extension of the FIA database focuses on invasive plants and assesses the occurrence of species that are a concern to forestry, but only for forested plots. Presence and absence information for the five species was extracted from this component of the FIA database. A species was recorded as present if one tree is found in an FIA plot (673 m^2); this is equivalent to a density of 15 invasive trees ha^{-1}. Thus in this chapter, invaded forest is defined as one or more

target species per plot. The FIA and GBIF data were considered independent since the former dataset is not part of GBIF.

For comparison of model prediction, county level data from Biota of North America Program (BONAP) was used (Kartesz, 2012; Figure 7-1). BONAP is a unified digital system for assessing North American biota. Its database includes an assessment for all vascular plants and vertebrate animals (native, naturalized and adventive) of North America and maintains the most current taxonomy, nomenclature and biogeographic data for all members of the biota.

7.3.4 Environmental variables

The composition of southeastern North American forest communities has been shown to be generally well correlated with climate variables (Swenson & Weiser 2010). Given the accessibility at a global and regional scale within the Bioclim database (www.worldclim.org), 19 bioclimatic variables were used at a thirty second resolution (approx 1 km x 1 km) along with elevation. This gave twenty variables derived from the monthly temperature, precipitation and elevation values to generate biologically meaningful parameters (Table 7-2).Climate variables often have high inter-correlation as they are based on precipitation and temperature and this can confound models and substantially increase processing time. Thus, Pearson's correlation was used for variable reduction to ensure all variables retained in the model were not non highly correlated variables, those that represent the extremes were selected (i.e. minimum temperature rather than mean temperature) as it has been suggested that these are more ecologically relevant to plant species (Zimmermann et al. 2009). As a result, environmental variables were reduced to elevation (ELEV) and eight weakly correlated climate variables: diurnal temperature range (BIO2), minimum temperature in the coldest month (BIO6), mean temperature of wettest quarter (BIO8), mean temperature of driest quarter (BIO9), annual precipitation (BIO12), precipitation in the driest month (BIO14), precipitation seasonality (BIO15) and precipitation of warmest quarter (BIO18).

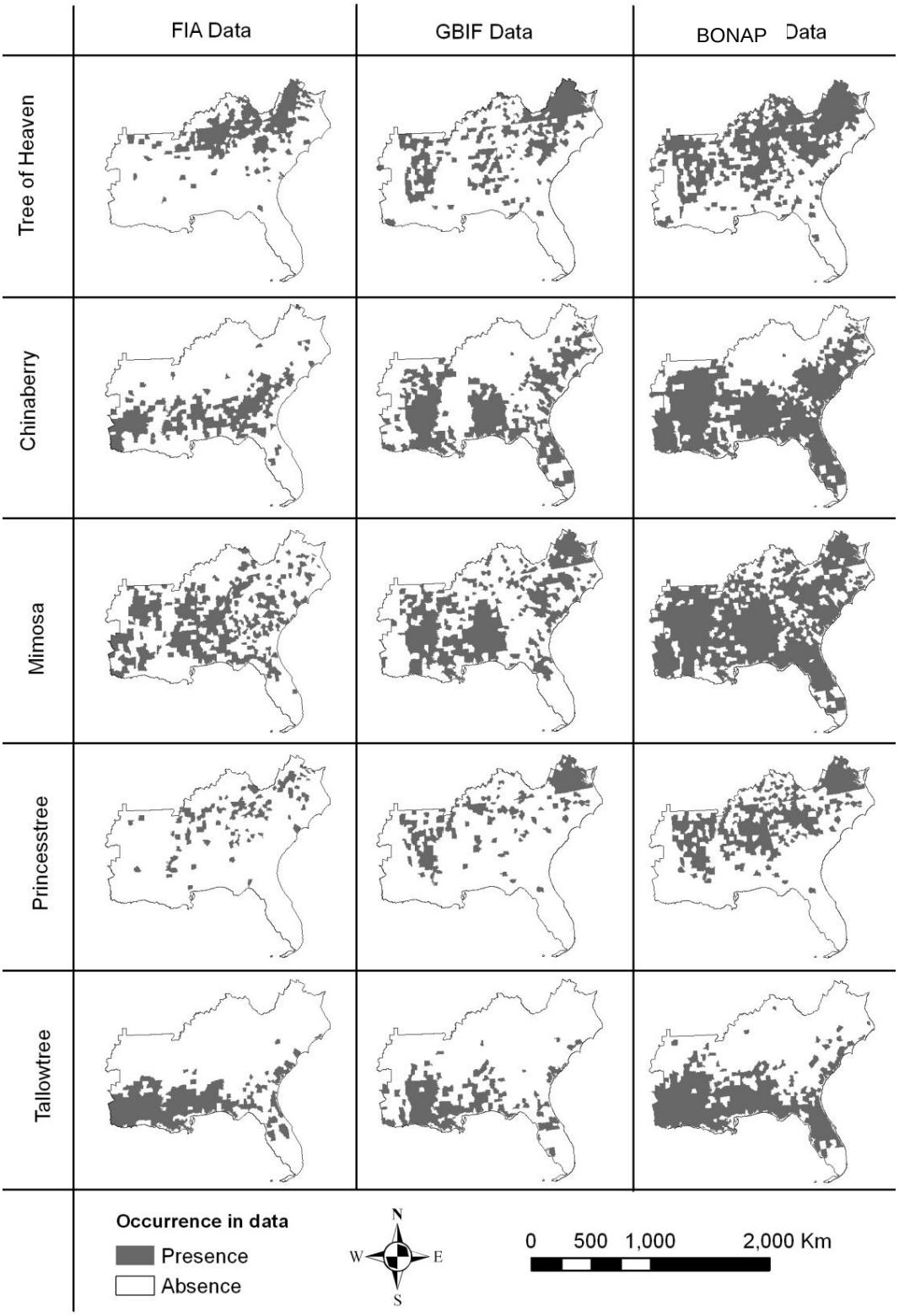

Figure 7-1: FIA and GBIF data (used to develop species distribution models) represented as county level occurrence, and BONAP data representing the most compressive county level occurrence information.

To test the similarity of environmental variables across datasets t-tests were conducted for each of the nine environmental variables with low correlations. To assess the if there was significant difference between overall background environmental characteristics in the southeastern region and sampled locations four sets of t-tests were undertaken that compare the southeastern background data (the forested region of the southeast as represented by 90x90 m grids) with: full FIA data set (9 t-tests), all GBIF sites in the southeastern region with occurrence of any of the five species (assuming if once species was found the others would have been too, if they were there, 9 t-tests), FIA presences locations (to represent a similar data set as GBIF, 9 t-tests), all GBIF sites (global) with occurrence of any of the five species (assuming if once species was found the others would have been too, if they were there, 9 t-tests). Two further set of t-test were done, broken down by species, to assess the difference between FIA occurrence sites and regional GBIF sites (five species, 45 t-tests), and the difference between regional and global GBIF sites (five species, 45 t-tests).We conducted 126 t-tests, thus to give an overall significance of 0.05 each individual test had a significance value of 0.0004.

Table 7-2: Summary of Bioclim variables and their means for the appropriate background data used in MaxEnt, low correlated variables that were retained for further analysis are shaded grey.

Code	Variable	Units/formula	Global		Regional	
			Mean	SD	Mean	SD
BIO1	Annual Mean Temperature	°C x 10	82	148	165	27
BIO2	Mean Diurnal Range	Mean of monthly (max temp - min temp) x10	112	31	128	9
BIO3	Isothermality	BIO2/BIO7x100	38	21	39	4
BIO4	Temperature Seasonality	SDx100	8321	5203	7181	1040
BIO5	Max Temperature of Warmest Month	°C x 10	259	102	323	16
BIO6	Min Temperature of Coldest Month	°C x 10	-91	199	-2	40
BIO7	Temperature Annual Range	BIO5-BIO6	350	138	325	34
BIO8	Mean Temperature of Wettest Quarter	°C x 10	153	106	180	68
BIO9	Mean Temperature of Driest Quarter	°C x 10	24	205	166	67
BIO10	Mean Temperature of Warmest Quarter	°C x 10	187	96	254	17
BIO11	Mean Temperature of Coldest Quarter	°C x 10	-25	204	69	39
BIO12	Annual Precipitation	mm	700	674	1304	149
BIO13	Precipitation of Wettest Month	mm	116	110	148	26
BIO14	Precipitation of Driest Month	mm	20	30	74	14
BIO15	Precipitation Seasonality	CV	61	34	20	10
BIO16	Precipitation of Wettest Quarter	mm	308	293	404	67
BIO17	Precipitation of Driest Quarter	mm	72	103	251	41
BIO18	Precipitation of Warmest Quarter	mm	208	187	350	83
BIO19	Precipitation of Coldest Quarter	mm	127	204	319	75
ELEV	Elevation	m	3555	2061	885	515

7.3.5 Species Distribution Models

SDMs can be constructed in a variety of ways and with a range of outputs. Most models predict species presence or absence, or just presence, to get geographic ranges, based on the environmental data associated with species occurrence records. SDMs can be developed using a variety of algorithms including statistical models (e.g. GLMs (Guisan et al. 1998; Randin et al. 2006), machine learning (Harrison et al. 2006; Phillips et al. 2006)), heuristic models (Beaumont & Hughes 2002) and combinatorial optimization (Fitzpatrick et al. 2007). There are two types of common outputs from these models: binary and continuous. Binary output defines each location as being within or outside the distribution and often requires a threshold to determine this boundary. Continuous results allow a less defined approach to identifying areas of high and low probability of occurrence.

We used the maximum entropy algorithm (MaxEnt) approach (Phillips et al. 2006) because we needed an algorithm that could use both presence/absence data and presence only data, and maximum entropy has consistently been among the top performing algorithms for SDM (Phillips et al. 2006; Elith et al. 2006; Ortega-Huerta & Peterson 2008). Maximum entropy is a machine learning technique that predicts species distributions using detailed geospatial data sets along with species occurrence information, and is conducted using the specialized package MaxEnt (Phillips et al. 2006). MaxEnt was used to develop probability envelopes for the occurrence of the five invasive trees in southeastern forests using distribution-environment relationships derived at the global (GBIF) and regional (FIA) data. Additionally, further analysis was undertaken at a regional scale using only regional GBIF records. This was to determine whether systematically acquired presence/absence data (FIA) and opportunistically acquired and presence only data (GBIF) from the same region gives different perspectives on distribution-environment relationships or whether the SDMs follow similar patterns. Models were derived using a manual backward selection method, where variables that had little or no impact on the model were removed, this was repeated ten times with different selections of training and test data. The key variables determining the occurrence of each species were identified by their percent contribution to the final model and with a jack-knife test on gain and influence on the area under the curve (AUC).

Four techniques were used to assess model reliability: the performance of test and training data, the omission rate, AUC and comparison with an external dataset. Occurrence data were randomly split with thirty percent in test and seventy percent in training datasets for the regional

models and run ten times with random selections. It has been shown that the area defined by the absences, pseudo-absences or background can have a major influence on model output (Phillips et al. 2009; Elith et al. 2010). Since GBIF only provides presence data, pseudo-absences were taken as random background points within the maximum latitude (+ 0.5 degrees) recorded in the GBIF data for each of the five species (Table 7-1), limiting the pseudo-absences to areas that the species could potentially invade. FIA data records both presence and absence, thus absence points were used to define the background. For the assessment of models all test data were limited to the southeastern region i.e. the global GBIF model used thirty percent of the data within the southeastern region as test data with all test data outside of the southeastern region dropped from the analysis.

The omission rate is the false negative or the proportion of sites where the species was present but the model predicted absence. To calculate this, a cut-off criterion is required to convert continuous model predictions to binary classifications. We used a threshold value that maximized the sum of sensitivity and specificity. This has the advantage of giving equal weight to the probability of success for both presence and absence (Manel et al. 2002). This is one of the most appropriate methods to derive a binary variable from continuous probabilities when species presence-absence distribution data are unbalanced (Jiménez-Valverde & Lobo 2006; Liu et al. 2005). This cut-off was also used for mapping of one run of the final selection of variables. AUC provides a single measure of model performance, independent of any particular choice of threshold but is sensitive to the method in which absences in the evaluation data are selected (Lobo et al. 2008). We used the following classes of AUC to assess model performance: 0.50 to 0.75 = fair, 0.75 to 0.92 = good, 0.92 to 0.97 = very good, and 0.97 to 1.00 = excellent (Hosmer & Lemeshow 2000). AUC is most applicable to data with true absences (Jiménez-Valverde 2011), thus it needs to be used with caution for datasets that do not have absence data. Other studies have shown the strength of using AUC as an assessment tool when predicting species one at a time over the same extent for all models, removing highly correlated variables and restricting the spatial distribution of all models to a similar environment, as was done in this study (Stohlgren et al. 2010). The proportion of agreement and spatial correlation between models were also assessed, with Pearson's correlation calculated between FIA model and global GBIF models and FIA models and regional GBIF models.

To address the value added by integrating SMDs from different datasets two approaches were taken; first an ensemble approach in which models were combined, and second by combining the

data before modelling. This was only done at the regional level. With the continuous outputs from the regional GBIF models and FIA models added together to give an ensemble map. This was reclassified to a binary map using the average maximized the sum of sensitivity and specificity. The combined data modelling approach was undertaken the same as with individual datasets. For comparison BONAP data were used, this is a county level occurrence dataset. All modelled data were converted to presence/absence for each county and omission rates were calculated between this and BONAP data.

7.4 Results

The results are reported in three parts; a comparison between models derived from the different datasets, the influence and relevance of the independent variables (environmental), and the combined modelling approaches.

7.4.1 Model Comparisons (GBIF, Regional GBIF and FIA)

The potential distribution of the five invasive tree species in the southeastern region was dependent on the source data used to make the predictions (Table 7-3, Figure 7-2). For each species, global GBIF SDMs presented less error in models as assessed by AUC and omission rate, but suggested between two to four times greater areas of potential distribution (51 – 99%) compared with models based on the regional FIA data (12 – 44%) and regional GBIF data (23 – 41%). Global GBIF models, as evaluated in the southeastern region, all had AUC above 0.92 and omission rates below 0.04, and when evaluated with BONAP data the omission rates were all below 0.01. However, they also predicted over 85 percent of the forest in the region as having the potential to be invaded by four of the five species. Regional GBIF models all had a much lower AUC (0.64 – 0.83) and higher omission rates (0.16 – 0.49), but they predicted less than 41 percent of the forest in the region as having the potential to be invaded by any of the species. When compared with the BONAP county level data omission rates dropped substantially for the regional GBIF data, with omission rates between 0.03 and 0.16. Regional FIA models had varying AUC and omission rates, generally better than the regional GBIF models and worse than the global GBIF models. Overall, FIA regional models had AUCs between 0.69 and 0.92, and omission rates between 0.13 and 0.35. FIA regional models and predicted less than 44 percent of the forest in the region having the potential to be invaded by any one of the species. In general the BONAP omission rates were similar to other FIA test omission rates. Silktree was the only

species that differed substantially, dropping from 0.35 to 0.15. Only one species, tallow tree, had FIA AUC over 0.9. Silktree had the poorest FIA model with an AUC only of 0.69 and an omission rate of 0.35.

Table 7-3: Accuracy assessment and proportion of forest with potential of invasion for the invasive trees of the southeastern United States (threshold, defined as maximum sensitivity plus specificity, test statistics only run on southeastern region for the global model).

Species	Model	N	MaxSS Threshold	Omission Rate			AUC		% of forest area potentially invasible
				train	test	BONAP	train	test	
Tree of Heaven	Global GBIF	2316	0.196	0.09	0.01	0.01	0.92	0.92	92
	Regional GBIF	302	0.415	0.30	0.34	0.13	0.77	0.76	34
	Regional FIA	855	0.260	0.09	0.13	0.16	0.89	0.88	32
	Combined Data	1157	0.363	0.18	0.12	0.14	0.85	0.88	33
	Combined Models	1157	0.335			0.15			33
Chinaberry	Global GBIF	887	0.308	0.22	0	0	0.88	0.99	85
	Regional GBIF	342	0.447	0.36	0.49	0.11	0.73	0.68	23
	Regional FIA	232	0.339	0.12	0.16	0.16	0.86	0.84	26
	Combined Data	574	0.395	0.18	0.10	0.06	0.76	0.83	36
	Combined Models	574	0.393			0.15			24
Silktree	Global GBIF	716	0.261	0.11	0.01	0.01	0.95	0.97	99
	Regional GBIF	403	0.457	0.27	0.37	0.16	0.70	0.64	41
	Regional FIA	681	0.444	0.27	0.35	0.15	0.72	0.69	44
	Combined Data	1084	0.487	0.34	0.34	0.10	0.66	0.67	46
	Combined Models	1084	0.451			0.17			41
Princess tree	Global GBIF	443	0.185	0.07	0.01	0	0.97	0.98	93
	Regional GBIF	213	0.262	0.13	0.36	0.06	0.84	0.74	35
	Regional FIA	484	0.369	0.14	0.13	0.07	0.82	0.82	40
	Combined Data	697	0.440	0.29	0.23	0.06	0.82	0.81	30
	Combined Models	697	0.315			0.07			36
Tallowtree	Global GBIF	310	0.132	0.07	0.04	0	0.97	0.98	51
	Regional GBIF	168	0.205	0.09	0.16	0.03	0.87	0.83	26
	Regional FIA	1093	0.233	0.17	0.13	0.13	0.92	0.92	12
	Combined Data	1251	0.185	0.10	0.07	0.07	0.90	0.92	19
	Combined Models	1251	0.219			0.06			19

All models had a much higher occurrence rate than the current distribution as measured by the percentage of FIA plots occupied (0.7 to 3.2 %). Regional GBIF and FIA models largely reflected the current distribution of the species, with similar numbers of counties predicted to be invaded as predicted by BONAP (Table 7-4) and had distinct distributions for each species. Conversely, for results of the global GBIF models, with the exception of tallowtree, the

distributions were remarkably similar to each other differing, only in the degree to which they penetrated northern Florida (Figure 7-2) and having much higher number of counties with occurrence than given by BONAP (Table 7-4).

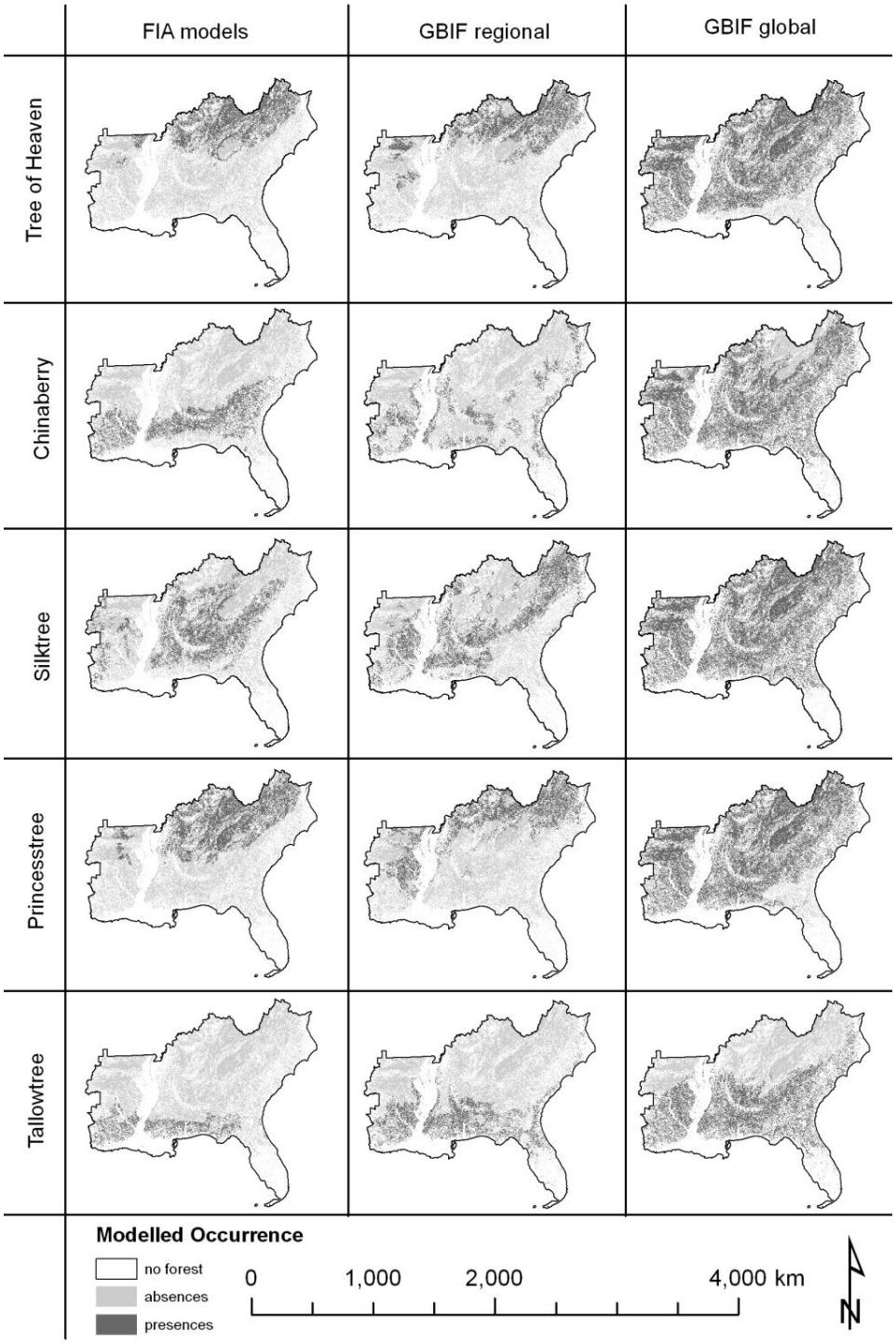

Figure 7-2: Potential invasion probability maps for five invasive trees in the forest of the southeastern region of the United States, developed from regional FIA data and GBIF

Table 7-4: Percentage of counties occupied by invasive trees, for input data and modelled data, shaded light grey if more than 5% lower than BONAP, shaded dark grey if more than 5% higher than BONAP.

	Tree of Heaven	Chinaberry	Silktree	Princesstree	Tallowtree
BONAP	47	50	66	34	29
Input FIA	23	20	30	11	18
Input GBIF	28	30	37	20	14
Modelled FIA	42	41	73	50	18
Modelled Regional GBIF	50	52	69	60	40
Modelled GBIF	92	95	98	92	65
Combined Data	45	56	80	54	32
Combined Models	45	44	66	53	31

Overall, occurrence agreement was much higher between the GBIF global and regional models than between eth GBIF regional models and FIA, however disagreement between models was also higher (Table 7-5). Tree of heaven had the most similar distribution between the regional GBIF models and FIA model, with 0.46 correlation and 74 percent spatial agreement in the model (Table 7-5). Silktree had the least agreement between regional models with only 46 percent spatial agreement and a correlation of 0.07 (Table 7-5).

Table 7-5: Spatial congruence between models (represented as a percentage) and correlation between models (represented as r)

	between GBIF global models and GBIF regional models					between GBIF regional global models and FIA regional models				
	Tree of Heaven	Chinaberry	Silktree	Princesstree	Tallowtree	Tree of Heaven	Chinaberry	Silktree	Princesstree	Tallowtree
Both models absences	8	15	1	8	49	59	61	43	49	73
Disagreement in models	60	59	55	52	39	16	30	54	27	16
Both models presences	32	26	44	40	12	25	9	3	24	11
Spatial correlation between models (r)	0.20	0.24	0.08	0.23	0.37	0.46	0.15	0.07	0.28	0.35

7.4.2 Environmental Variables

We found very strong similarity between the environmental variables for the regional background and the full FIA data. The regional GBIF data and occurrence only FIA data had some similarity and the global GBIF data had very little (Table 7-6). Between global GBIF and regional GBIF 16 out of 45 pairs were not significantly different (Table 7-7). There was no clear

pattern across species or variables. However, mean diurnal temperature range differed significantly between global and regional GBIF data sets for all species. , Six out of nine variables did not differ between the global and regional GBIF data sets for princesstree (Table 7-7). There was strong similarity between GBIF species locations and FIA species location, with 26 out of 45 pairs of independent variables having no significant difference (Table 7-8). Minimum temperature of coldest month, annual precipitation and precipitation of the warmest quarter were not significantly different between sites of occurrence in global GBIF and FIA data for four of the five species (Table 7-8). Sites with princesstree differed in only one variable, elevation, and sites with tallowtreee differed in two, precipitation of the driest month and precipitation seasonality (Table 7-8).

Table 7-6: Mean environmental data for combined species datasets, t-test with southeastern region background data, pairs that are not significantly different (t-test) are shaded in grey.

	Regional Background	FIA	Regional GBIF	FIA Occurrence	Global GBIF
Elevation	175	173	137	163	262
Mean Diurnal Range	130	130	127	129	108
Min Temperature of Coldest Month	-5	-6	-5	-1	3
Mean Temperature of Wettest Quarter	175	174	182	179	172
Mean Temperature of Driest Quarter	163	162	156	165	116
Annual Precipitation	1314	1315	1296	1322	1109
Precipitation of Driest Month	75	75	76	78	53
Precipitation Seasonality	20	21	18	17	28
Precipitation of Warmest Quarter	350	348	344	345	321

Table 7-7: Mean environmental data for global GBIF and regional GBIF, pairs that are not significantly different (t-test) are shaded in grey.

	Tree of Heaven		Chinaberry		Silktree		Princesstree		Tallowtree	
	region	global	region	global	region	global	region	global	region	global
Elevation	216	273	82	304	139	260	176	173	54	192
Mean Diurnal Range	128	103	126	114	128	117	127	113	122	102
Min Temperature of Coldest Month	-29	-21	14	71	-7	-11	-28	-27	30	62
Mean Temperature of Wettest Quarter	180	144	189	216	176	187	179	172	192	219
Mean Temperature of Driest Quarter	121	82	178	188	160	121	134	94	196	186
Annual Precipitation	1217	827	1336	1413	1304	1348	1226	1153	1426	1730
Precipitation of Driest Month	76	47	74	49	77	65	77	67	78	70
Precipitation Seasonality	15	21	21	49	18	28	15	18	22	38
Precipitation of Warmest Quarter	316	217	370	429	339	408	310	307	394	604

Table 7-8: Mean environmental data for FIA occurrence and regional GBIF occurrence, pairs that are not significantly different (t-test) are shaded in grey.

	Tree of Heaven		Chinaberry		Silktree		Princesstree		Tallowtree	
	region	FIA	region	FIA	region	FIA	region	FIA	region	FIA
Elevation	216	312	82	94	139	169	176	296	54	44
Mean Diurnal Range	128	130	126	132	128	131	127	129	122	124
Min Temperature of Coldest Month	-29	-42	14	16	-7	-7	-28	-35	30	33
Mean Temperature of Wettest Quarter	180	180	189	164	176	150	179	168	192	205
Mean Temperature of Driest Quarter	121	112	178	193	160	182	134	135	196	190
Annual Precipitation	1217	1222	1336	1284	1304	1332	1226	1266	1426	1422
Precipitation of Driest Month	76	78	74	68	77	75	77	79	78	82
Precipitation Seasonality	15	13	21	21	18	19	15	14	22	18
Precipitation of Warmest Quarter	316	317	370	346	339	333	310	319	394	381

Although six variables were identified as important in one or more of the global GBIF models, two (minimum temperature in the coldest month and annual precipitation) were common to all models and on average accounted for 62 percent of the explainable variation (Table 7-9). In contrast, eight variables were included in one or more regional FIA model with only one variable shared among all models and accounting for 53 percent of the explainable variation (minimum temperature in the coldest month; Table 7-9). The regional GBIF models used one variable not used by any of the other datasets (mean diurnal range) and this was the second most dominant variable explaining an average of 24 percent of the explainable variation in all regional GBIF models. The most dominant variable in the regional GBIF models was minimum temperature in the coldest month, which accounted for an average of 36 percent of the explainable variation for all models. For the majority of models, the same species had variables in the same direction or were non-linear in one of the models, with 41 to 90 percent explainable variation shared between models (Table 7-9). The only species for which this was not true was chinaberry, which was also the only species whose range extends through the tropics.

Table 7-9: Contribution of dominant variables based on ten replicates (- negative, + positive, ∩ or U for quadratic relationship).

	Tree of Heaven			Chinaberry			Silktree			Princesstree			Tallowtree				
	GBIF	GBIF - Reg	FIA	GBIF	GBIF - Reg	FIA	GBIF	GBIF - Reg	FIA	GBIF	GBIF - Reg	FIA	GBIF	GBIF - Reg	FIA	# of models	Average
Elevation			∩10	-16	-33	-11			∩32		-9		-15	-13		8	9
Mean Diurnal Range		-10			-34			-31			-19			-25		5	8
Min Temperature of Coldest Month	∩33	-28	-55	26	33	∩64	∩21	∩21	∩17	∩29	∩27	-60	8	∩42	∩74	15	35
Mean Temperature of Wettest Quarter									∩16		∩7					2	2
Mean Temperature of Driest Quarter		-11							13			-9		11		4	3
Annual Precipitation	∩18	-28	-9	23		-14	61	17	∩12	∩37	-23		55	9		12	20
Precipitation of Driest Month				∩28				∩23	∩10				22			4	6
Precipitation Seasonality	-49	-23	-26				-18	-8		-34	-15	-20			∩26	9	15
Precipitation of Warmest Quarter				7		-11						∩11				3	2

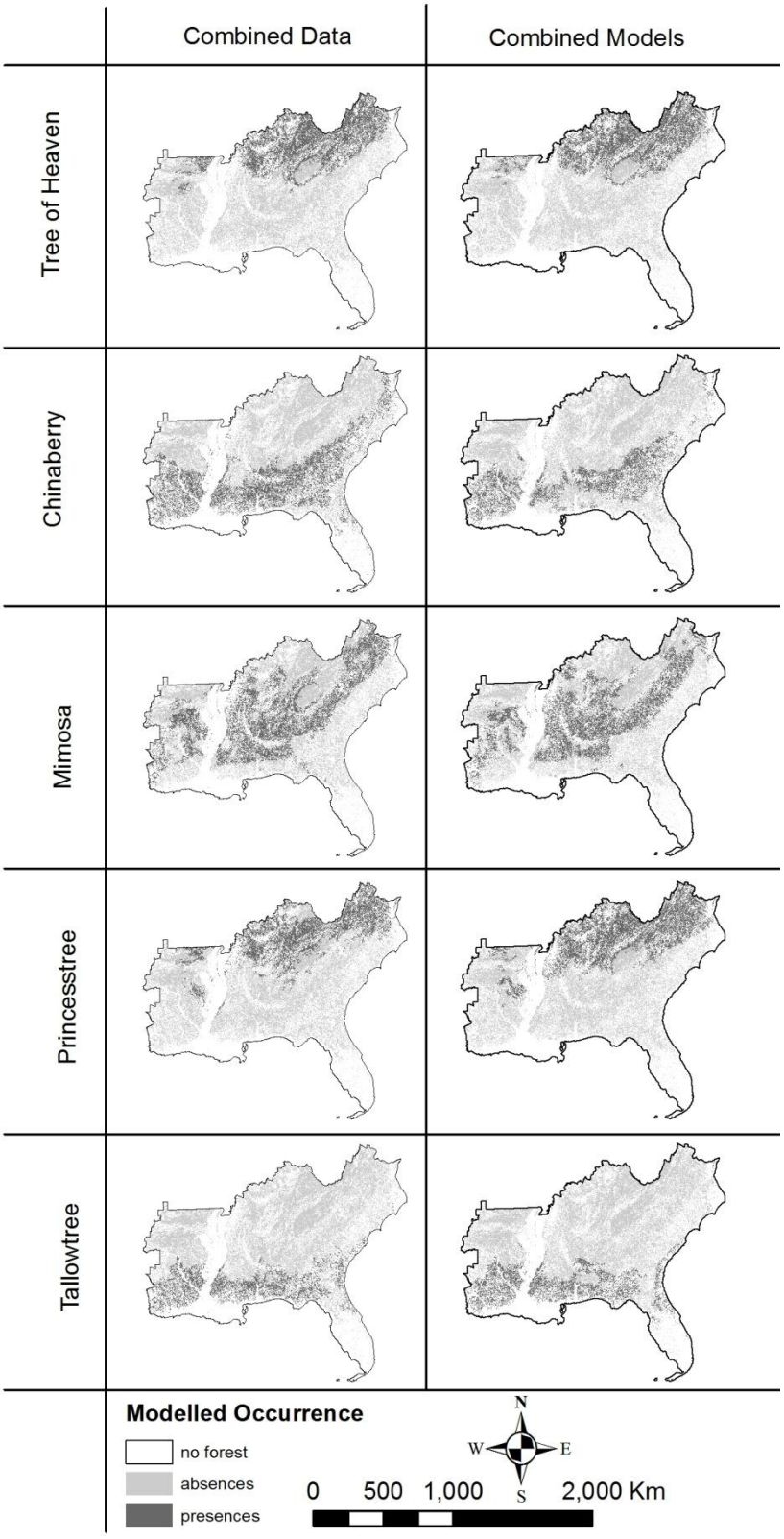

Figure 7-3: Potential invasion probability maps for five invasive trees in the forest of the southeastern region of the United States, developed combined regional FIA and GBIF data

7.4.3 Combined Models

Assessment of the combined data were done in two ways, combining regional GBIF data with FIA data and remodelling (combined data approach), and combining the models developed independently for regional GBIF data with FIA data (combined models approach). Overall the combined data models had lower test omission rates and higher test AUC rates, with princesstree test omission rate being an exception. The combined data approach gave better statistics than the combined model approach, with the combined data approach, on average having a better BONAP omission of 0.03 (Table 7-3). The combined data approach predicted occurrence in more counties, second only to the global GBIF models (Table 7-4). The geographic distribution (Figure 7-3) was similar for both approaches.

7.5 Discussion

The goal of this study was to assess the impact of data in SDMs, focussing on the influence of scale (global vs regional), data type (opportunistically acquired vs systematically acquired), consistency of environmental variables (independent) across scales, and the value of combining datasets. These issues are fundamental to all species distribution modelling but of particular interest to invasive species, which often have expanding distributions and for which limited information may be available in areas of interest. Overall we showed large variation in modelled distribution based on the spatial scale of the data used, as well as some variation based on data type, with similarity in environmental variables.

There are a number of studies in which invasive species are modelled with global data to assess the probability of invasion at a regional level (Kikillus et al. 2010) or by matching native climatic conditions with potential ranges (Rodda et al. 2009, Thuiller et al. 2005). Other studies take the approach of modelling within a region based on regional data to identify areas most at risk or already invaded (Lemke et al 2010). Few studies have examined the impact of regional opportunistically acquired data as compared with global opportunistically acquired data (Sanchez-Fernandez et al 2011). This is the approach we have taken. We found models built on global data covering larger geographical ranges gave agreement with Sanchez-Fernandez and others (2011). The GBIF data has a global range, including both native and invaded habitats around the world and in many physiographical regions. Given the limitations of only using climate and elevation data, we suggest that the global GBIF models represent a broader range of conditions that may not be limited by the same biotic and anthropogenic constraints found in the

southeastern region. Thus global scale model may give an indication of potential invasion, but once a species is invading the regional distribution will give a more refined model with a narrower distribution. This is contrary to Sanchez-Fernandez and others (2011) who suggested species data from regional inventories, may provide an incomplete description of the environmental limits of most species resulting in a biased description of species' niches. We, however, think that by only integrating climatic variables and using a global dataset one will overestimate the potential distribution of species through the limited influence of regional characteristics on global models.

With the tremendous growth in data availability, both for species occurrence and regional geospatial data and user friendly species distribution modelling tools there has been a proliferation in SDMs in the last five years (Franklin 2009). There have been numerous studies comparing SDM methods (Elith et al. 2006) but few comparing data types and sources. In this study we have compared models developed from opportunistically acquired (GBIF) and systematically acquired (FIA). For three of five species the models of spatial distribution produced using opportunistically acquired and systematically acquired data were similar, resulting in spatial agreement of occurrence at over 40% of sites. Of the five species, three had a slightly larger distribution predicted by models constructed with systematically acquired data suggesting that models built on opportunistically acquired data will not always overestimate. The close agreement between models suggests strength in these models and the high AUC and low omission rates confirm this. One species, silktree, had very weak spatial agreement between the models (5% of occurrence points). Silktree also had very low AUC (<0.70) for both regional models suggesting that we have not successfully modelled the distribution of this species at the regional level, and that its distribution is likely being driven by something we failed to identify. The value of using different datasets can be seen in this evaluation of models by identifying three strong models and one very weak model. By combining models it is also possible to show areas of high probability of occurrence (all models agree) and moderate probability of occurrence (models disagree). This has the potential to offer land managers, particularly those that work at a regional level, a broader picture of the future landscape. Overall the models derived using systematically acquired data (FIA) were stronger than the models derived from opportunistically acquired data regional (GBIF) models in terms of validation statistics. In general there were more FIA points than GBIF regional points but this was not true for all species, thus the number of data points cannot be the only component driving the more accurate FIA models. The FIA

data points were measured within metres of their true location, but the GBIF data is at an unknown level of accuracy; however, in some cases the GBIF points seem to be based on county centroids (potentially 100's of kilometres from where the species was found). This could be the reason for decreased accuracy found in the GBIF models. The FIA data also has absence data, and in the case of an invading species not at equilibrium, this could be more detrimental to the model than productive. However , this influence is limited by the maximum entropy algorithm by which the sampling points (absences and presence) are considered as background value that define the range of sample environments rather than true absences.

Many studies have found large scale environmental factors can produce strong SDMS (Franklin 2009). Overall we found similar results, with the environmental variables used in this study useful in predicting the species distribution, with 20 out of 25 models considered good (test AUC > 0.75, and test omission rate < 0.25). For the majority of models, the same species had variables in the same direction or were non-linear in one of the models, with 41 to 90 percent contribution of variables shared between models. One species, tallowtree, had very strong models across both scale and datasets (AUC > 0.82 and omission rate < 0.14). This suggests that tallowtree distribution is strongly driven by climate given that we used climate-based variables and models derived from systematically and opportunistically acquired data can equally represent its distribution. Tallowtree had a higher current distribution, than any of the species as assessed by the USFS (Miller et al. 2012). For the other species there was some variation across datasets. At a global level (GBIF) for all species models were strong, but the regional models they were weaker. This suggests that, overall, species distribution is climate driven but at a regional level there are other variables influencing the distribution. Temperature in the coldest month was the only variable used in all 15 models, with an average contribution of 35 percent. It has been suggested that SDMs should use variables related to heat, light, moisture and nutrients (Mackey 1993). In this study we did not integrate any nutrient-related variables as they are not yet well represented geospatially on a global scale and the addition of this information may add value to SDMs.

A number of studies have used an ensemble approach, using a number of modelling techniques (Grenouillet et al 2010, Capinha and Anastacio 2011). In this case we took an ensemble approach with models build with different datasets. The strongest regional models were in general the ones derived from the combined FIA and regional GBIF data, with

princesstree the only exception. This suggests that though the quality of data is important, quantity is fundamental.

Understanding the environmental limits of a species is essential to predicting its potential distribution (Araujo & Guisan, 2006). However, the estimation of these limits is not free of error, even if a broad scale intensively sampled data set is used. SDMs are one tool that can be used to assist in identifying areas of greatest risk of invasion and in developing management strategies, but they are not the only tool and need to be used in context of the data that was used in their development. Although this work was undertaken to understand the application of data on invasive species, not at equilibrium, we would suggest that the same would hold true for species in their native ranges and this should be examined to determine the applicability of global data to regional SDMs in general.

7.6 Literature Cited

Andersen, M.C., Adams, H., Hope, B. & Powell, M. (2004) Risk analysis for invasive species: general framework and research needs. Risk analysis, 24, 893-900.

Araújo, M. B. and Guisan, A. (2006), Five (or so) challenges for species distribution modelling. Journal of Biogeography, 33: 1677–1688

Beaumont, L.J., Gallagher, R.V., Downey, P.O., Thuiller, W., Leishman, M.R. & Hughes, L. (2009) Modelling the impact of *Hieracium* spp. on protected areas in Australia under future climates. Ecography, 32, 757-764.

Brown, K.A., Spector, S. & Wu, W. (2008) Multi-scale analysis of species introductions: combining landscape and demographic models to improve management decisions about non-native species. Journal of Applied Ecology, 45, 1639-1648.

Bystriakova, N. & Schneider, H. (2011) Evolution of the climatic niche in scaly tree ferns (Cyatheaceae , Polypodiopsida). Botanical Journal of the Linnean Society, 165, 1-19.

Edwards Jr., T., Cutler, D., Zimmermann, N., Geiser, L. & Moisen, G. (2006) Effects of sample survey design on the accuracy of classification tree models in species distribution models. Ecological Modelling, 199, 132-141.

Elith, J. & Leathwick, J.R. (2009) Species distribution models: Ecological explanation and prediction across space and time. Annual Review of Ecology Evolution and Systematics, 40, 677-697.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. Overton, J., Townsend Peterson, A., Phillips, S.J., Richardson, K., Scachetti-pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography, 29, 129-151.

Elith, J., Kearney, M. & Phillips, S. (2010) The art of modelling range-shifting species. Methods in Ecology and Evolution, 1, 330-342.

Fitzpatrick, M.C., Weltzin, J.F., Sanders, N.J. & Dunn, R.R. (2007) The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? Global Ecology and Biogeography, 16, 24-33.

Flemons, P., Guralnick, R., Krieger, J., Ranipeta, A. & Neufeld, D. (2007) A web-based GIS tool for exploring the world's biodiversity: The Global Biodiversity Information Facility Mapping and Analysis Portal Application (GBIF-MAPA). Ecological Informatics, 2, 49-60.

Franklin, J. (2009) Mapping species distributions: spatial inference and prediction, 3rd ed. Cambridge University Press, New York.

Gallagher, R.V., Beaumont, L.J., Hughes, L. & Leishman, M.R. (2010) Evidence for climatic niche and biome shifts between native and novel ranges in plant species introduced to Australia. Journal of Ecology, 98, 790-799.

Gallien, L., Münkemüller, T., Albert, C.H., Boulangeat, I. & Thuiller, W. (2010) Predicting potential distributions of invasive species: where to go from here? Diversity and Distributions, 16, 331-342.

Graham C.H., Ferrier S., Huettman F., Moritz C. & Peterson A.T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology & Evolution, 19, 497-503.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. Ecological Modelling, 135, 147-186.

Guisan, A., Theurillat, J.P. & Kienast, F. (1998) Predicting the potential distribution of plant species in an alpine environment. Journal of Vegetation Science, 9, 65-74.

Harrison, P., Berry, P., Butt, N. & New, M. (2006) Modelling climate change impacts on species' distributions at the European scale: implications for conservation policy. Environmental Science Policy, 9, 116-128.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25, 1965-1978.

Hosmer, D.W. & Lemeshow, S. (2000) Applied Logistic Regression. Wiley-Interscience, New York.

Hulme, P.E. & Weser, C. (2011) Mixed messages from multiple information sources on invasive species: a case of too much of a good thing? Diversity and Distributions, 7, 1152–1160.

Jiménez-Valverde, A. & Lobo, J.M. (2006) The ghost of unbalanced species distribution data in geographical model predictions. Diversity and Distributions, 12, 521-524.

Jiménez-Valverde, A. (2011) Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. Global Ecology and Biogeography, 21, 498-507.

Kartesz, J.T. (2012) Floristic Synthesis of North America, Version 1.0. Biota of North America Program (BONAP). (in press)

Kruijer, H., Raes, N. & Stech, M. (2010) Modelling the distribution of the moss species *Hypopterygium tamarisci* (Hypopterygiaceae, Bryophyta) in Central and South America. Nova Hedwigia, 91, 399-420.

Lemke, D., Hulme, P.E., Brown, J.A. & Tadesse, W. (2011) Distribution modelling of Japanese honeysuckle (*Lonicera japonica*) invasion in the Cumberland Plateau and Mountain Region, USA. Forest Ecology and Management, 262, 139-149.

Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R. (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography, 28, 385-393.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography, 17, 145-151.

Mackey, B.G. (1993) A spatial analysis of the environment relations of rainforest structural types. Journal of Biogeography, 20, 303-336.

Manel, S., Williams, H.C. & Ormerod, S.J. (2002) Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology, 38, 921-931.

Miller, J.H., Chambliss, E.B. & Loewenstein, N.J. (2010) A field guide for the identification of invasive plants in southern forests. Gen. Tech. Rep. SRS–119. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 126 p.

Miller, J.H., Lemke, D. & Coulston, J. (2012) The Southern Forest Futures Project, Chapter 15. The Invasion of Southern Forests by Nonnative Plants: Current and Future Occupation with Impacts, Management Strategies, and Mitigation Approaches. In Press.

Nori, J., Urbina-Cardona, J.N., Loyola, R.D., Lescano, J.N. & Leynaud, G.C. (2011) Climate change and american bullfrog invasion: What could we expect in South America? PLoS ONE, 6, e25718.

Ortega-Huerta, M.A. & Peterson, A.T. (2008) Modeling ecological niches and predicting geographic distributions : a test of six presence-only methods. Revista Mexicana De Biodiversidad, 79, 205- 216.

Peterson, A.T. & Vieglais, D.A. (2001) Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. BioScience, 51, 363.

Peterson, A.T. (2003) Predicting the geography of species invasions via ecological niche modeling. Quarterly Review of Biology, 78, 419-433.

Phillips, S., Anderson, R. & Schapire, R. (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190, 231-259.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications, 19, 181-197.

Pimentel, D., Zuniga, R. & Morrison, D. (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecological Economics, 52, 273-288.

PLANTS (2011) PLANTS National Plants Database. National Plant Data Center, Baton Rouge. http://plants.usda.gov.

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006) Are niche-based species distribution models transferable in space? Journal of Biogeography, 33, 1689-1703.

Reddy S. & Dávalos L.M. (2003). Geographical sampling bias and its implications for conservation priorities in Africa. Journal Of Biogeography, 30, 1719-1727.

Ricciardi, A. (2007) Are modern biological invasions an unprecedented form of global change? Conservation Biology, 21, 329-36.

Sánchez-Fernández D., Lobo J.M., Abellán P., Ribera I. & Millán, A. 2008. Bias in freshwater biodiversity sampling: the case of Iberian water beetles. Diversity and Distributions, 14: 754-762.

Smith, W.B., Miles, P.D., Perry, C.H. & Pugh, S.A. (2009) Forest Resources of the United States 2007. Gen. Tech. Rep. WO-78. Washington, DC.

Soberón J.M., Llorente J.B. & Oñate L. (2000). The use of specimen-label databases for conservation purposes: an example using Mexican Papilionid and Pierid butterflies. Biodiversity and Conservation, 9, 1441-1466.

Stohlgren, T.J., Ma, P., Kumar, S., Rocca, M., Morisette, J.T., Jarnevich, C.S. & Benson, N. (2010) Ensemble habitat mapping of invasive plant species. Risk Analysis, 30, 224–235.

Swenson, N.G. & Weiser, M.D. (2010) Plant geography upon the basis of functional traits: an example from eastern North American trees. Ecology, 91, 2234-41.

USDA/FS (2007) The Forest Inventory and Analysis Database: Database Description and Users Guide Version 3.0. U.S. Department of Agriculture, Forest Service, Washington Office.

Vitousek, P.M., D'Antonio, C.M., Loope, L.L., Rejmanek, M. & Westbrooks, R. (1997) Introduced species: A significant component of human-caused global change. New Zealand Journal of Ecology, 21, 1-16.

Wang, H.-H., Grant, W.E., Swannack, T.M., Gan, J., Rogers, W.E., Koralewski, Tomasz E. Miller, J.H. & Taylor Jr, J.W. (2011) Predicted range expansion of Chinese tallow tree (*Triadica sebifera*) in forestlands of the southeastern United States. Diversity and Distributions, 17, 552-565.

Wear, D.N. & Greis, J.G. (2002) Southern forest resource assessment: summary of findings. Journal of Forestry, 100, 6–14.

Wear, D.N. & Greis, J.G. (2012) Summary report the southern forest futures project. In Press.

Wilcove, D.S., Rothstein, D., Dubow, J., Phillips, A. & Losos, E. (1998) Quantifying threats to imperiled species in the United States. BioScience, 48, 607.

Williams P.H., Margules C.R. & Hilbert D.W. (2002). Data requirements and data sources for biodiversity priority area selection. Journal Of Biosciences, 27, 327-338.

Zimmermann, N.E., Edwards, T.C., Moisen, G.G., Frescino, T.S. & Blackard, J. A. (2007) Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. The Journal of Applied Ecology, 44, 1057-1067.

Zimmermann, N.E., Yoccoz, N.G., Edwards, T.C., Meier, E.S., Thuiller, W., Guisan, A., Schmatz, D.R. & Pearman, P.B. (2009) Climatic extremes improve predictions of spatial patterns of tree species. Proceedings of the National Academy of Sciences of the United States of America, 106, 19723-19728.

CHAPTER 8.

Summary

In this thesis, we assessed and modelled invasion of alien plant species in the forest of the southeastern United States. There are over 380 recognized invasive plants in southeastern forests and grasslands with 53 ranked as high-to-medium risk to natural communities (Miller et al. 2010). We focused on ten of these: Chinese lespedeza, tall fescue, Japanese honeysuckle, Chinese privet, autumn olive, princesstree, silktree, chinaberry, tree of heaven, tallowtree (Table 8-1). We assess them at differing scales (Figure 8-1), locally (Chapter 2 and 3), eco-regionally (Chapter 4 and 5) and regionally (Chapters 6 and 7), using anthropogenic and environmental variables from field based, remotely sensed and digital datasets, and apply both parametric and non-parametric modelling approaches, thus addressing ecological and statistical issues of species distribution models (SDMs).

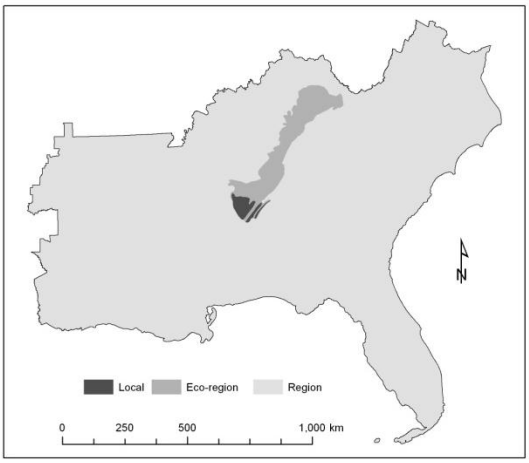

Figure 8-1: Scale of study: local, eco-region and region.

Table 8-1: Comparison of contribution of variable groups (environmental, anthropogenic or remotely sensed) across scales for species distribution models of the ten focal species (L=local, E=eco-region, R=regional, 0 to 25% lowercase, 25 to 50% uppercase, 50% + uppercase bold)

		Environmental	Anthropogenic	Remotely Sensed
Chinese lespedeza	*Lespedeza cuneata*	l	L	**L**
Tall fescue	*Lolium arundinaceum*	**E**	E	e
Japanese honeysuckle	*Lonicera japonica*	l, **E**	L, e	**L,** e
Chinese privet	*Ligustrum sinense*	l, **E**	L, E	**L,** e
Autumn olive	*Elaeagnus umbellata*	l	L	**L**
Princesstree	*Paulownia tomentosa*	l, **R**	L, r	**L**
Silktree	*Albizia julibrissin*	**E,** R	E, **R**	
Chinaberry	*Melia azedarach*	**R**	R	
Tree of heaven	*Ailanthus altissima*	**R**	r	
Tallowtree	*Triadica sebifera*	**R**	r	

SDM's provide description of the relationship between species distribution and environmental parameters and are often used to make predictions for unsampled locations (Franklin 2009). Environmental characteristics play an important role in determining a site's vulnerability to invasion and on-going establishment. Once an alien species occurs in a new region, climate and the broad scale land cover have strong influences on the extent of invasion. The sites at which the species occurs within the invasion range are often driven by small scale characteristics, including micro climate, nutrient availability and forest community characteristics (Franklin 2009). Human activities have the most influence on invasion progression through dispersal (movement and introduction rate; Gelbard & Belnap 2003, Avon et al. 2010) and disturbance of the landscape (increased resource availability; Buckley et al. 2007, Pysek 2002, Davis et al. 2000). At an eco-region and regional scale, environmental characteristics dominated (>50%) all but one species distribution model (silktree at the regional scale; Table 8-1). At the eco-regional scale, elevation was the dominant variable, and at a regional scale, minimum temperature was the dominant variable. The eco-regional level can be considered an intermediate scale, in which climatic conditions have similarities. Some of the smaller scale characteristics influence distribution but, due to the size of the area in this study, it is not practical to build models using data at a finer scale. At regional or global scales similar elevation does not necessarily result in similar climatic characteristics. Thus, while elevation comes out as a strong independent variable at the eco-regional scale, it is acting as a surrogate for temperature variation with higher areas being cooler, and for topographic characteristics, with the low areas being flatter. Overall this trend confirms the utility of matching the climate ranges of native species with the range of potential invasion (Peterson et al. 2003) for regional assessments. It also suggests that climate change will strongly influence the distribution and that variation in climate should be integrated into models.

Anthropogenic information added some value to the models; however they were rarely the dominant variables (Table 8-1). Roads and land use (proportion of forest or distance to forest) were the most useful anthropogenic variables. In all final (combined anthropogenic and environmental) models evaluated, only six times did any one anthropogenic variable represent more than 25 percent of the model and four of these were at the local scale. At the regional and eco-regional level, roads had a greater than 25 percent contribution to the silktree models while at a local level, distance to forest and distance to roads contributed more than 25 percent to the models for sawtooth oak, Japanese honeysuckle and privet. All the species considered in this

study are ones that are considered at high-to-medium risk of invading natural communities. In most case they have passed the introduction phase and are in the establishment phase. We suggest that given the observed dominance of environmental variables these species have been broadly established throughout the area. In terms of model application and method transferability all of this information, environmental and anthropogenic, is readily available for North America and for much of the world, making this level of landscape level modelling very practical.

At a local scale, in contrast to the regional and eco-regional scale, no model was dominated by environmental characteristics (Table 8-1). We used a third group of variables, remotely sensed variables, and these could relate to either environmental or anthropogenic characteristics. Remotely sensed variables are a direct measurement of the environmental characteristics at a given point in time as measured through light reflectance of the land cover and may relate to primary productivity, disturbance, light or nutrient availability. The only models they were useful in were at the local scale (also evaluated at an eco-regional level, but not a regional level), where they represented approximately 70% percent of the variables contributing to all local models. The variation of climatic variables is limited at a local scale, thus they add little to models at this scale. The preference of remotely sensed variables over other environmental variables that had dominated the eco-region and regional models related to resolution and habitat availability. At a local level we were able to use remotely sensed variables that were at a much finer resolution than are practical at an eco-region or regional level due to technical and data constraints. Within the local level we focused on species that were present at the local level, thus were in a climatic environment that was suited to their growth. Where the species grow within this climatically suitable environment is likely to be driven by small scale resource available hence the dominance of variables that relate to resources availability (light, disturbance, etc).

Four species were assessed at multiple scales, Japanese honeysuckle, Chinese privet, princesstree and silktree. Both Japanese honeysuckle and Chinese privet were assessed at local and eco-regional scales, with remotely sensed variables dominating the models at a local scale and environmental at a regional scale. Princestree showed a similar relationship with remotely sensed variables dominating at a local scale and environmental at a regional scale. This is consistent with much of the literature, that although specific local occurrence is often determined by local environmental heterogeneity, disturbance and resources availability the broader potential distribution is predominantly driven by climatic conditions (Franklin 2009). This relationship did not hold true across eco-regional to regional scales. For only one species,

silktree, were the models dominated at the eco-regional scale by environmental variables, and at the regional scale by anthropogenic variables. These models were not strong, suggesting we were missing some component of what is driving the distribution of this species.

At regional and global scales, we found models built on global data predicted much larger geographical ranges than those built on regional data. This agrees with other recent studies (Sanchez-Fernandez et al. 2011). The global models used data including both native and invaded habitats around the world and in many physiographical regions, representing a broader range of conditions that may not be limited by the same biotic and anthropogenic constraints found in the southeastern United States. Sanchez-Fernandez and others (2011) suggested species data from regional inventories may provide an incomplete description of the environmental limits of most species resulting in a biased description of species' niches. Only integrating climatic variables and using a global dataset will overestimate the potential distribution of species due to the limited influence of regional characteristics in global models.

From a statistical perspective, in building models of complex systems from limited data assessment of the models is imperative. Along with the tradition assessment tools, such as omission rate and AUC, we also compared models across model types, and data types, evaluating their strength through agreement in spatial distribution, and direction of the influence of independent variables. The application of two modelling approaches, logistic regression and MaxEnt, showed inclusion of similar variables with the similar impacts. The geographical agreement between models adds confidence. Overall the logistic regression models had slightly better omission rates and the MaxEnt models had better AUC's. THe logistic regression models also often predicted larger geographical areas of occurrences when the threshold of maximum sensitivity plus specificity was used, thus the lower omission rates are likely related to the less stringent model that predicts a larger area. Alternatively using the same model but different datasets can give similar information. Again overall we found inclusion of similar variables with similar impacts. We also found there was more value in the intensive surveyed data but this was less dramatic than in using the defined areas of interest to select the data for models.

Overall, the potential for invasion at the local scale is best modelled through remotely sensed and anthropogenic variables, whereas eco-regional and regional invasion potential by environmental variables. An integrative modelling approach, using more than one modelling technique adds greater strength to conclusions. If just using one approach, MaxEnt may be

slightly more preferred as it on the whole gives a more defined predicted area of occurrence and generally had higher AUC values.

Predicting the future distribution of invasive plants is pivotal to planning effective forest management but is challenged by the fact that expanding populations are rarely at equilibrium with their environment due to progression of invasion and ongoing changes in the invaded landscape (land cover change, climate change). Through this work we have identified both species and areas of most concern, suggesting that invasive plants in the forest of the southeastern regional will continue to expand in distribution. Models such as those developed by this research can be used as tools for landscape management, forest stand assessment or long term forest monitoring programs. One of the greatest benefits of large-scale GIS models is that they can outline the main characteristics of species distribution areas and be used to predict environmental favourability in regions where their distribution is less documented (Barbosa et al. 2009). They can also be integrated into forest management decision support systems (Ducheyne et al. 2006) and assist in developing long term management plans. There needs to be more work in technology transfer and making the outputs of these models informative for managers. To further increase the application of SDM integrating potential scenarios with the likelihood of the scenarios into the resultant SDM would add value. A recent paper by Elith et al. (2010) suggested that it is essential to think about the biology of the species, integrate the fitted models and reduce as much as possible unwanted effects. This is the approach taken in this study within the limitations of the knowledge of these species and data constraints. However this needs to be further assessed with more study into species response to environmental constraints.

SDMs can be a useful tool in assessing potential impacts of invasive species and the impact future conditions may have on the distribution of invasives, but is currently limited by our knowledge of the species responses and by uncertainty about future conditions. Such models are however a useful start for improving our knowledge and, with some refinement, will become a key management tool. Fundamentally in the application of SDMs the selection of scale, independent and dependent data and model type need to be based on the question the SDM is being used answer.

8.1 Literature Cited

Avon, C., Bergès, L., Dumas, Y. & Dupouey, J.L. (2010) Does the effect of forest roads extend a few meters or more into the adjacent forest? A study on understory plant diversity in managed oak stands. Forest Ecology and Management, 259, 1546-1555.

Barbosa, A.M., Real, R. & Mario Vargas, J. (2009) Transferability of environmental favourability models in geographic space: the case of the Iberian desman (*Galemys pyrenaicus*) in Portugal and Spain. Ecological Modelling, 220, 747–754.

Buckley, Y.M., Bolker, B.M. & Rees, M. (2007) Disturbance, invasion and re-invasion: managing the weed-shaped hole in disturbed ecosystems. Ecology Letters, 10, 809-17.

Davis, M. A., Grime, J.P. & Thompson, K. (2000) Fluctuating resources in plant communities: a general theory of invasibility. Journal of Ecology, 88, 528-534.

Ducheyne, E.I., De Wulf, R.R. & De Baets, B. (2006) A spatial approach to forest-management optimization: linking GIS and multiple objective genetic algorithms. International Journal of Geographical Information Science, 20, 917–928.

Elith, J., Kearney, M. & Phillips, S. (2010) The art of modelling range-shifting species. Methods in Ecology and Evolution, 1, 330-342.

Franklin, J. (2009) Mapping species distributions: spatial inference and prediction, 3rd ed. Cambridge University Press, New York.

Gelbard, J.L. & Belnap, J. (2003) Roads as conduits for exotic plant invasions in a semiarid landscape. Conservation Biology, 17, 420-432.

Miller, J.H., Chambliss, E.B. & Loewenstein, N.J. (2010) A field guide for the identification of invasive plants in southern forests. Gen. Tech. Rep. SRS–119. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 126 p.

Peterson, A.T., Papes, M. & Kluza, D.A. (2003) Predicting the potential invasive distributions of four alien plant species in North America. Weed Science, 51, 863-868.

Pysek, P. (2002) Patterns of invasion in temperate nature reserves. Biological Conservation, 104, 13-24.

Sánchez-Fernández D., Lobo J.M., Abellán P., Ribera I. & Millán, A. (2008) Bias in freshwater biodiversity sampling: the case of Iberian water beetles. Diversity and Distributions, 14: 754-762.

CHAPTER 9.

Summary and Correlation Tables

Table A9-1: Pearson correlations among ground variables in grey indicating highly correlated variables (r>0.71), Chapter 2.

	%Rock	%BS	%NVP	%DWD	%Shale	%LL	%GL	%Litter	Litter	Humus
%Under	0.06	-0.14	-0.14	-0.20	-0.13	-0.48	0.38	-0.35	-0.18	-0.11
%Rock		0.13	0.04	-0.15	-0.06	-0.20	0.01	-0.25	-0.14	-0.11
%BS			-0.02	-0.19	0.26	-0.45	0.10	-0.50	-0.48	-0.30
%NVP				-0.12	0.10	0.08	-0.11	0.03	-0.05	-0.01
%DWD					-0.16	0.27	-0.32	0.12	0.25	0.21
%Shale						-0.29	-0.03	-0.38	-0.33	-0.19
%LL							-0.59	0.85	0.62	0.43
%GL								-0.08	-0.28	-0.21
%Litter									0.58	0.39
Litter										0.64

Table A9-2: Pearson correlations among forest structure variables in grey indicating highly correlated variables (r>0.71), Chapter 2.

	Shan	Simp	Hrich	Canopy cover	% upper	%mid	DEN	SDEN	MDEN	LDEN	BA	PDEN
Rich	0.90	0.71	0.98	0.54	0.34	0.59	0.54	0.51	0.39	0.18	0.26	0.05
Shan		0.93	0.87	0.50	0.32	0.56	0.41	0.41	0.28	0.15	0.19	-0.01
Simp			0.68	0.42	0.27	0.46	0.31	0.32	0.20	0.11	0.14	-0.01
Hrich				0.48	0.30	0.56	0.47	0.46	0.32	0.18	0.23	-0.06
Orich				0.37	0.21	0.49	0.39	0.41	0.23	0.14	0.17	-0.10
Canopy cover					0.68	0.53	0.57	0.39	0.59	0.47	0.61	0.39
%upper						0.08	0.28	0.11	0.34	0.61	0.72	0.16
%mid							0.58	0.57	0.42	0.02	0.06	0.29
DEN								0.92	0.83	0.06	0.23	0.80
SDEN									0.57	-0.05	0.01	0.69
MDEN										0.00	0.31	0.76
LDEN											0.87	-0.03
BA												0.14

Table A9-2 continued: Pearson correlations among forest structure variables in grey indicating highly correlated variables (r>0.71), Chapter 2.

	SPDEN	MPDEN	LPDEN	PBA	HDEN	SHDEN	MHDEN	LHDEN	HBA	HHSDEN	HHSBA	Age
Rich	-0.02	0.1	0.05	0.09	0.84	0.81	0.79	0.5	0.58	0.58	0.71	0.3
Shan	-0.04	0.03	0.03	0.05	0.71	0.68	0.67	0.44	0.51	0.48	0.56	0.33
Simp	-0.03	0.01	0.02	0.03	0.54	0.52	0.5	0.34	0.39	0.34	0.4	0.29
Hrich	-0.1	0.01	0.03	0.05	0.86	0.83	0.82	0.54	0.61	0.6	0.74	0.29
Canopy	0.16	0.47	0.41	0.53	0.42	0.41	0.38	0.26	0.31	0.25	0.3	0.41
%upper	-0.06	0.27	0.58	0.68	0.24	0.24	0.22	0.16	0.2	0.21	0.17	0.38
%mid	0.29	0.23	-0.09	-0.07	0.56	0.53	0.54	0.41	0.45	0.38	0.41	0.18
DEN	0.69	0.69	0.02	0.17	0.56	0.57	0.46	0.14	0.2	0.18	0.37	0.11
SDEN	0.76	0.41	-0.09	-0.05	0.58	0.61	0.47	0.14	0.2	0.18	0.38	0.02
MDEN	0.43	0.93	-0.02	0.29	0.34	0.35	0.31	0.07	0.11	0.1	0.22	0.15
LDEN	-0.16	-0.05	0.96	0.82	0.13	0.11	0.14	0.24	0.23	0.17	0.12	0.34
BA	-0.13	0.26	0.84	0.96	0.19	0.17	0.18	0.2	0.22	0.17	0.15	0.33
PDEN	0.88	0.83	0.02	0.2	-0.05	-0.01	-0.09	-0.18	-0.18	-0.15	-0.12	-0.05
SPDEN		0.49	-0.12	-0.08	-0.07	-0.06	-0.08	-0.15	-0.15	-0.12	-0.12	-0.15
MPDEN			-0.01	0.31	0	0.04	-0.07	-0.16	-0.16	-0.13	-0.08	0.05
LPDEN				0.87	-0.01	0	-0.02	-0.03	-0.03	-0.04	-0.03	0.28
PBA					0.01	0.03	-0.03	-0.08	-0.07	-0.07	-0.04	0.25
HDEN						0.98	0.91	0.5	0.6	0.52	0.78	0.25
SHDEN							0.82	0.39	0.49	0.43	0.73	0.22
MHDEN								0.59	0.71	0.6	0.79	0.27
LHDEN									0.96	0.78	0.54	0.25
HBA										0.82	0.64	0.3
HHSDEN											0.67	0.28
HHSBA												0.2

Table A9-3: Pearson correlations among soil variables in grey indicating highly correlated variables (r>0.71), Chapter 2.

	P	K	Na	Mg	Ca	Fe	Zn	Cu	Mn	Ca:Mg	NH4	NO3	%C	%N	%S	C:N	CEC	%sat
pH	0.03	0.12	0.08	0.31	0.7	-0.17	-0.01	0.15	0.17	0.46	-0.04	0.21	0.04	0.11	-0.03	0.05	-0.1	0.74
P		0.49	0.15	0.24	0.31	0.37	0.31	0.36	-0.03	0.02	0.29	0.24	0.3	0.3	0.46	0.13	0.21	0.3
K			0.41	0.39	0.34	0.3	0.23	0.24	0.26	-0.11	0.22	0.25	0.15	0.28	0.18	-0.08	0.49	0.26
Na				0.38	0.25	0.21	0.27	0.36	0.17	-0.12	-0.05	0.25	0.28	0.31	0.27	0.14	0.22	0.14
Mg					0.57	0.37	0.47	0.62	0.16	-0.25	-0.14	0.23	0.25	0.37	0.19	-0.03	0.49	0.54
Ca						0.2	0.38	0.56	0.16	0.39	0.02	0.37	0.33	0.42	0.27	0.12	0.13	0.9
Fe							0.49	0.58	0	-0.25	0.01	0.1	0.58	0.55	0.34	0.38	0.38	0.17
Zn								0.68	0.15	-0.17	-0.04	0.17	0.32	0.4	0.23	0.03	0.25	0.32
Cu									0.08	-0.16	-0.15	0.29	0.51	0.59	0.44	0.16	0.28	0.49
Mn										-0.07	0.15	0.09	0.11	0.17	0.13	-0.01	0.12	0.11
Ca:Mg											0.18	0.11	-0.1	-0.13	-0.06	0.09	-0.26	0.43
NH4												0.35	0.02	0	0.05	0.11	0.2	-0.09
NO3													0.23	0.28	0.16	0.09	-0.02	0.34
%C														0.89	0.67	0.52	0.14	0.28
%N															0.61	0.2	0.25	0.33
%S																0.23	0.06	0.23
C:N																	-0.08	0.17
CEC																		0

Table A9-4: Pearson correlations among selected soil feature, ground, forest structure variables, grey indicating highly correlated variables (r>0.71), Chapter 2.

	Mg	Zn	Mn	NH4	%N	CEC	%Under	%DWD	%Shale	%GL	%Litter	Humus	Simp	Canopy cover	%Mid	HDEN
Na	0.32	0.31	0.1	-0.15	0.24	0.18	-0.12	0.13	-0.07	-0.1	0.2	0.10	0.03	0.24	0.02	0.06
Mg		0.43	0.11	-0.05	0.29	0.57	0.05	-0.09	-0.17	-0.02	0.27	-0.06	-0.02	0.22	-0.08	-0.15
Zn			0.2	-0.03	0.38	0.27	0.14	-0.05	0.02	0.16	0.02	0.02	-0.06	-0.05	-0.11	-0.14
Mn				0.19	0.15	0.11	0.03	0.3	-0.25	0.05	0.12	0.12	0.21	0.07	0.03	0.15
NH4					0.04	0.21	0.07	0.07	-0.05	0.02	-0.02	0.07	0.16	0.01	0.15	0.3
%N						0.25	-0.05	0.15	0.04	-0.09	0.23	0.14	0.06	0.19	-0.02	-0.01
CEC							0	0.03	-0.07	-0.18	0.23	0.13	0.17	0.31	0.14	0.07
% Under								-0.18	-0.13	0.38	-0.34	-0.11	-0.19	-0.44	-0.33	-0.33
% DWD									-0.16	-0.32	0.12	0.21	0.26	0.34	0.13	0.19
% Shale										-0.03	-0.38	-0.19	-0.1	-0.29	-0.15	-0.16
%GL											-0.08	-0.21	-0.31	-0.54	-0.28	-0.22
% Litter												0.39	0.15	0.68	0.3	0.27
Humus													0.23	0.37	0.35	0.38
Simp														0.42	0.47	0.53
Canopy cover															0.53	0.42
%Mid																0.56

Table A9-5: Pearson correlations among landscape variables in grey indicating highly correlated variables (r>0.71), Chapter 3.

	North	Curve	Slope	Solar	Water	Open	Dist for	Roads	ff90	ff00	ff06	NDVI 09	NDVI 87	NDVI 92	NDVI 98	NDVI 04	NDVI 11
Age	0	-0.02	-0.03	-0.03	0.11	0.12	0.11	0.22	0.3	0.32	0.36	-0.03	0.45	0.11	-0.3	-0.1	-0.02
North		0.04	0.04	-0.04	-0.01	-0.01	-0.02	-0.02	0.03	-0	-0	0	0.07	0	0.01	0.01	0.01
Curve			0.01	-0.09	0.01	-0.01	-0.01	0	-0.1	0	0.01	0	-0.01	0	-0.03	0	-0.03
Slope				-0.73	-0.13	-0.26	-0.26	-0.09	-0	0.23	0.21	0.03	-0.07	-0.01	0.05	0.33	0.33
Hill					0.07	0.15	0.16	0.07	0.03	-0.2	-0.1	0	0.05	0.01	-0.07	-0.17	-0.19
Water						-0.04	0.01	-0.16	0.08	-0.1	-0.1	0.01	0.19	-0.09	0.03	0.06	0.2
Open							0.89	0.18	0.01	-0.6	-0.7	-0.13	0.25	0.25	-0.17	-0.71	-0.46
Dist for								0.18	0.11	-0.5	-0.6	-0.09	0.27	0.21	-0.09	-0.65	-0.45
Roads									-0.2	-0	0.02	0.07	0.15	0.2	-0.04	-0.24	-0.22
ff90										0.03	-0	-0.06	0.35	0.26	0.21	0.21	0.1
ff00											0.9	0.22	-0.28	0.12	0.39	0.69	0.46
ff06												0.19	-0.34	0.09	0.35	0.62	0.4
NDVI09													-0.04	-0.05	0.01	0.3	0.42
NDVI87														0.71	0.02	-0.18	0.01
NDVI92															0.45	-0.03	-0.02
NDVI98																0.27	0.18
NDVI04																	0.71

Table A9-6: Summary statistics for landscape variables, Chapter 4 and 5.

	Ave	SD	Min	Max	Skew	Kurt	Trans	Ave	SD	Min	Max	Skew	Kurt
DI75	9.66	1.28	5.2	27.23	2.12	22.27	Log	2.26	0.13	1.65	3.3	0.26	4.75
DI90	-0.11	2	-3.22	14.71	2.51	10.06	Log	1.26	0.43	-0.24	2.93	0.53	0.8
DI00	0.4	2.4	-3.6	17.5	2.64	9.41	Log	1.37	0.43	-0.97	30.6	0.71	1.65
DI90-75	-9.28	2.61	-28.1	8.59	1.53	7.58							
DI00-90	0.49	2.5	-13	18.84	1.45	9.53							
NDVI75	0.49	0.16	-0.41	0.77	-1.16	1.64							
NDVI90	0.56	0.1	-0.11	0.76	-1.39	4.34							
NDVI00	0.42	0.16	-0.31	0.73	-0.61	0.42							
NDVI90-75	-0.07	0.22	-0.93	0.69	0.2	0.53							
NDVI00-90	-0.14	0.17	-0.86	0.56	-0.28	0.88							
CENSUS	24	48	3	1152	12.4	223	log	2.78	0.79	1.09	7.05	0.79	1.91
RD_DIST	365	366	0	2665	1.92	4.78	sqrt	16.39	7.58	1	50.5	0.2	0.04
RD_DEN	1.4	1.1	0	9.9	1.95	7.63	log	0.8	0.43	0	2.39	0.07	0.03
MRD_DIST	5531	4742	0	25529	1.28	1.5							
RES ALL	0.79	0.4	0	1	-1.45	0.13							
RES100	0.24	0.42	0	1	1.25	-0.44							
RES500	0.72	0.45	0	1	-0.97	-1.05							
MINT	26.55	3.36	21	35	0.29	-1.39							
MAXT	86.7	2.67	77	91	-0.07	-0.47							
AVET	57	2.9	49	63	0.39	-1.06							
RAIN	54	5	41	69	-0.44	-0.61							

Table A9-6 continued: Summary statistics, Chapter 4 and 5.

	Ave	SD	Min	Max	Skew	Kurt	Trans	Ave	SD	Min	Max	Skew	Kurt
NORTH	1.01	0.71	0	2	0	-1.2							
EAST	0.99	0.71	0	2	0.02	-1.2							
NORTHNESS	0	0.18	-0.55	0.6	0.11	0.54							
EASTNESS	0	0.15	-0.7	0.64	0.05	0.74							
SLOPE	12	8.7	0	44.3	0.67	-0.4							
HILL	238	16	153	254	-1.8	3.38							
CURV	0	0.44	-1.8	2.36	0.12	2.15							
DEM	372	163	80.4	1117	0.71	0.8							
FC100	46.8	62.5	-100	200	0.61	-0.49							
FC500	42.29	45.4	-81.35	192.1	0.65	-0.162							
F00 100	0.8	0.27	0.04	1	-1.4	0.84							
F00 500	0.75	0.22	0.01	1	-0.96	0.11							
FARM100	0.1	0.2	0	1.3	1.9	2.9							
FARM500	0.2	0.2	0	0.9	1.3	1.1							
LULC90	Category												
LULC00	Category												
RIV DIS	325	282	0	2554	1.68	6.09	sqrt	16.4	7.57	1	50.6	0.2	0.04
RIV_DEN	1	0.5	0	3.6	0.7	1							
WATER100	0.06	0.23	0	1	3.81	12.5							
WATER500	0.32	0.47	0	1	0.77	-1.4							

Table A9-7: Pearson's correlation for DI and NDVI, Chapter 4 and 5.

	Log DI 90	Log DI 75	DI 00-90	DI 90-75	NDVI 00	NDVI 90	NDVI 75	NDVI 00-90	NDVI 90-75
Log DI00	0.40	0.08	0.59	0.83	-0.35	-0.13	-0.08	-0.26	-0.01
Log DI90		0.20	-0.41	0.23	-0.11	-0.48	-0.11	0.18	0.00
Log DI75			-0.05	-0.38	-0.11	-0.29	-0.50	0.07	0.29
DI 00-90				0.65	-0.28	0.24	0.00	-0.42	-0.21
DI 90-75					-0.34	-0.02	0.16	-0.32	-0.37
NDVI 00						0.30	0.12	0.81	0.68
NDVI 90							0.24	-0.33	0.05
NDVI 75								-0.04	-0.66
NDVI 00-90									0.64

Table A9-8: Person correlation of anthropogenic variables, Chapter 4 and 5.

	Sqrt Rd dist	Log Rd den	Mrd dist	Res500	Res100	Res_all
Log census	0.09	0.38	-0.33	0.20	0.14	0.18
Sqrt RD dist		0.01	-0.05	-0.01	-0.07	-0.02
Log RD den			-0.20	0.56	0.35	0.48
MRD dist				-0.06	-0.05	-0.06
Res500					0.34	0.81
Res100						0.28

Table A9-9: Pearson's correlation for environmental variables, Chapter 4 and 5.

	north	northness	eastness	slope	hill	curv	DEM
east	-0.04	-0.05	0.82	0.00	0.39	-0.01	0.02
north		0.82	-0.05	0.05	-0.53	0.01	-0.06
northness			-0.07	0.03	-0.64	0.03	-0.06
eastness				0.00	0.48	-0.01	0.02
slope					-0.62	-0.01	0.24
hill						-0.03	-0.11
curv							0.05

Table A9-10: Pearson's correlation of climate variables, Chapter 4 and 5.

	maxT	aveT	Rain
minT	0.75	0.93	0.65
maxT		0.87	0.31
aveT			0.54

Table A9-11: Person's correlations for Land use variables, Chapter 4 and 5.

	Fc100	F00_500	F00_100	Farm500	Farm100
Fc500	0.78	0.09	0.11	-0.05	-0.07
Fc100		-0.00	0.21	0.02	-0.16
F00_500			0.70	-0.88	-0.62
F00_100				-0.62	-0.87
Farm500					0.69

Table A9-12: Persons correlations for water variables, Chapter 4 and 5.

	riv_den	Water500	Water100
Sqrt riv_dis	-0.49	-0.12	-0.22
riv_den		0.23	0.17
Water500			0.36

Table A9-13: Correlations of regional landscape variables, Chapter 6.

	BIO2	BIO6	BIO8	BIO9	BIO12	BIO14	BIO15	BIO18	DEM	North	Slope	Censu	RD	MRD	CITY	FOR	FARM	GRASS	PINE	RES	FORC	PINEC
BIO6	-0.27																					
BIO8	-0.3	0.31																				
BIO9	0.06	0.41	-0.31																			
BIO12	-0.06	0.3	-0.36	0.47																		
BIO14	0.02	-0.34	-0.42	0.14	0.57																	
BIO15	-0.23	0.7	0.39	0.11	0.03	-0.66																
BIO18	-0.3	0.64	0.46	-0.03	0.37	-0.14	0.64															
DEM	0.06	-0.65	-0.32	-0.4	0	0.35	-0.4	-0.22														
North	0	0	0	0	0	0	0	0	0													
Slope	0.07	-0.46	-0.2	-0.27	0	0.22	-0.27	-0.13	0.66	0												
Census	-0.11	0.02	0.03	-0.04	-0.05	0	-0.01	0.02	0	-0.01	-0.03											
RD	-0.12	0.09	0.01	0.02	0.07	-0.03	0.11	0.07	0.02	0	0.06	-0.09										
MRD	0.05	0.03	-0.09	0.1	0.14	0.04	0.03	-0.05	-0.04	0	0.02	-0.16	0.08									
CITY	0.29	-0.16	-0.05	-0.12	0.06	0.02	-0.02	-0.04	0.12	0	0.15	-0.22	0.04	0.23								
FOR	0.27	-0.46	-0.28	-0.19	-0.04	0.19	-0.33	-0.27	0.44	0	0.4	-0.11	-0.03	0.06	0.16							
FARM	-0.02	-0.1	0.01	-0.02	-0.15	-0.07	-0.05	-0.15	-0.05	0	-0.14	0.01	-0.15	-0.1	-0.12	-0.44						
GRASS	0.06	0.11	0.18	-0.08	-0.14	-0.18	0.15	0.15	-0.08	0	-0.07	-0.02	-0.06	-0.07	-0.01	-0.2	-0.04					
PINE	0.1	0.31	0.06	0.14	0.12	-0.05	0.12	0.17	-0.28	0	-0.22	-0.09	-0.06	0.06	0.02	0.37	-0.32	-0.05				
RES	-0.07	0.08	0.04	0.01	-0.02	-0.03	0.06	0.06	-0.02	-0.01	-0.07	0.5	-0.29	-0.15	-0.16	-0.21	0.03	0.01	-0.09			
FORC	-0.03	-0.06	0.05	-0.11	-0.1	-0.12	0.08	0.02	0.04	0	0.04	-0.04	0.1	0	0.02	0.31	0	-0.29	0.18	-0.12		
PINEC	0.04	-0.02	-0.02	0.02	-0.06	-0.05	-0.05	-0.13	-0.06	0	-0.04	-0.04	0.02	0.04	0.04	0.26	-0.05	-0.25	0.45	-0.07	0.55	
RIV	-0.09	0.27	0.22	0.01	0.01	-0.21	0.35	0.33	-0.13	0	-0.08	-0.01	0.04	0	0	-0.1	-0.06	0.14	0.11	0.04	0.04	-0.02

Table A9-14: Correlations of regional input variables

	BIO01	BIO02	BIO03	BIO04	BIO05	BIO06	BIO07	BIO08	BIO09	BIO10	BIO11	BIO12	BIO13	BIO14	BIO15	BIO16	BIO17	BIO18	BIO19
ELEV	0	0	0	0	0	0	0	-0.01	0	0	0	-0.01	-0.01	0	0	-0.01	0	-0.01	0
BIO01		0.52	0.82	-0.82	0.88	0.95	-0.72	0.79	0.92	0.91	0.96	0.36	0.44	0.04	0.37	0.43	0.07	0.21	0.24
BIO02			0.39	-0.22	0.71	0.36	0	0.53	0.44	0.62	0.43	-0.25	-0.1	-0.38	0.51	-0.12	-0.37	-0.21	-0.24
BIO03				-0.89	0.61	0.89	-0.83	0.63	0.81	0.66	0.89	0.56	0.58	0.22	0.28	0.58	0.25	0.35	0.43
BIO04					-0.51	-0.94	0.97	-0.5	-0.86	-0.59	-0.93	-0.55	-0.57	-0.24	-0.19	-0.57	-0.27	-0.33	-0.4
BIO05						0.77	-0.36	0.84	0.79	0.99	0.8	0.11	0.23	-0.15	0.43	0.21	-0.13	0.02	0.05
BIO06							-0.88	0.7	0.95	0.83	1	0.48	0.52	0.16	0.27	0.52	0.19	0.27	0.35
BIO07								-0.39	-0.78	-0.46	-0.85	-0.61	-0.59	-0.34	-0.08	-0.59	-0.37	-0.39	-0.47
BIO08									0.61	0.86	0.73	0.25	0.37	-0.07	0.45	0.35	-0.05	0.25	0.07
BIO09										0.83	0.95	0.37	0.41	0.1	0.26	0.4	0.13	0.14	0.3
BIO10											0.85	0.19	0.3	-0.08	0.41	0.28	-0.06	0.09	0.11
BIO11												0.45	0.51	0.12	0.32	0.5	0.15	0.26	0.31
BIO12													0.9	0.7	-0.17	0.92	0.74	0.79	0.75
BIO13														0.39	0.14	0.99	0.42	0.74	0.58
BIO14															-0.52	0.42	0.99	0.55	0.67
BIO15																0.1	-0.52	-0.1	-0.27
BIO16																	0.46	0.76	0.61
BIO17																		0.57	0.7
BIO18																			0.37

Table A9-15: Correlations of global input variables.

	ELEV	BIO01	BIO02	BIO03	BIO04	BIO05	BIO06	BIO07	BIO08	BIO09	BIO10	BIO11	BIO12	BIO13	BIO14	BIO15	BIO16	BIO17	BIO18
BIO01	-0.75																		
BIO02	0.06	-0.16																	
BIO03	-0.33	0.71	0.18																
BIO04	0.36	-0.8	0.22	-0.91															
BIO05	-0.81	0.79	0.14	0.3	-0.29														
BIO06	-0.65	0.97	-0.27	0.79	-0.91	0.62													
BIO07	0.34	-0.75	0.42	-0.81	0.97	-0.19	-0.89												
BIO08	-0.32	0.27	-0.3	0.19	-0.32	0.1	0.31	-0.34											
BIO09	-0.4	0.48	0.06	0.25	-0.24	0.5	0.41	-0.22	-0.31										
BIO10	-0.86	0.93	-0.08	0.45	-0.52	0.95	0.81	-0.47	0.19	0.52									
BIO11	-0.66	0.98	-0.18	0.81	-0.9	0.66	0.99	-0.86	0.3	0.41	0.83								
BIO12	0	0.27	-0.06	0.31	-0.34	0.04	0.3	-0.36	-0.36	0.47	0.16	0.3							
BIO13	-0.24	0.6	-0.17	0.62	-0.69	0.24	0.65	-0.67	0.02	0.33	0.42	0.65	0.71						
BIO14	0.35	-0.39	0.02	-0.27	0.27	-0.39	-0.34	0.2	-0.42	0.14	-0.38	-0.36	0.57	-0.01					
BIO15	-0.4	0.67	-0.23	0.63	-0.7	0.37	0.7	-0.66	0.39	0.11	0.51	0.71	0.03	0.68	-0.66				
BIO16	-0.26	0.61	-0.23	0.65	-0.73	0.21	0.68	-0.72	0.08	0.34	0.41	0.68	0.7	0.97	-0.02	0.71			
BIO17	0.29	-0.24	0.04	-0.19	0.18	-0.25	-0.22	0.14	-0.45	0.24	-0.24	-0.23	0.72	0.13	0.9	-0.61	0.1		
BIO18	-0.22	0.5	-0.3	0.7	-0.8	-0.03	0.64	-0.82	0.46	-0.03	0.22	0.62	0.37	0.71	-0.14	0.64	0.78	-0.06	
BIO19	-0.04	0.14	0.17	0.14	-0.08	0.09	0.12	-0.1	-0.59	0.61	0.14	0.13	0.82	0.33	0.62	-0.31	0.31	0.73	-0.02

Table A9-16: Summary of all independent variables used in Dawn Lemke thesis: Alien plants and their invasion of the forested landscape of the southeastern United States.

Code	Variable	Unit	Sources	Resolution	Chapter
pH	pH		field	2 samples of 1400mm^2 in 405m^2	ch2
P	Phosphorus	ppm	field	2 samples of 1400mm^2 in 405m^2	ch2
K	Potassium	ppm	field	2 samples of 1400mm^2 in 405m^2	ch2
Na	Sodium	ppm	field	2 samples of 1400mm^2 in 405m^2	ch2
Mg	Magnesium	ppm	field	2 samples of 1400mm^2 in 405m^2	ch2
Ca	Calcium	ppm	field	2 samples of 1400mm^2 in 405m^2	ch2
Fe	Iron	ppm	field	2 samples of 1400mm^2 in 405m^2	ch2
Zn	Zinc	ppm	field	2 samples of 1400mm^2 in 405m^2	ch2
Cu	Copper	ppm	field	2 samples of 1400mm^2 in 405m^2	ch2
Mn	Manganese	ppm	field	2 samples of 1400mm^2 in 405m^2	ch2
Ca:Mg	Calcium Magnesium ratio		field	2 samples of 1400mm^2 in 405m^2	ch2
NH4	Ammonium	ppm	field	2 samples of 1400mm^2 in 405m^2	ch2
NO3	Nitrate	ppm	field	2 samples of 1400mm^2 in 405m^2	ch2
%C	% Carbon	%	field	2 samples of 1400mm^2 in 405m^2	ch2
%N	% Nitrogen	%	field	2 samples of 1400mm^2 in 405m^2	ch2
%S	% Sulphur	%	field	2 samples of 1400mm^2 in 405m^2	ch2
C:N	Carbon Nitrogen ratio		field	2 samples of 1400mm^2 in 405m^2	ch2
CEC	cation exchange capacity		field	2 samples of 1400mm^2 in 405m^2	ch2
%sat	% base saturation	%	field	2 samples of 1400mm^2 in 405m^2	ch2
%Under	% understory	%	field	2 samples of 10.2m^2 in 405m^2	ch2
%Rock	% Rock	%	field	2 samples of 10.2m^2 in 405m^2	ch2
%BS	% Bare Soil	%	field	2 samples of 10.2m^2 in 405m^2	ch2
%NVP	% non-vascular plants	%	field	2 samples of 10.2m^2 in 405m^2	ch2
%DWD	DWD	%	field	2 samples of 10.2m^2 in 405m^2	ch2
%Shale	% Shale	%	field	2 samples of 10.2m^2 in 405m^2	ch2
%LL	% Leaf Litter	%	field	2 samples of 10.2m^2 in 405m^2	ch2
%GL	% Grass Litter	%	field	2 samples of 10.2m^2 in 405m^2	ch2
%Litter	%Total Litter	%	field	2 samples of 10.2m^2 in 405m^2	ch2
Humus	Humus depth	cm	field	2 samples of 10.2m^2 in 405m^2	ch2
Litter	Litter Depth	cm	field	2 samples of 10.2m^2 in 405m^2	ch2

Table A9-16 Continued: Summary of all independent variables used in Dawn Lemke thesis: Alien plants and their invasion of the forested landscape of the southeastern United States.

Code	Variable	Unit	Sources	Resolution	Chapter
Rich	Richness		calculated from field	$405m^2$	ch2
Shan	Shannon		calculated from field	$405m^2$	ch2
Simp	Simpson's Evenness		calculated from field	$405m^2$	ch2
Hrich	Hardwood Richness		calculated from field	$405m^2$	ch2
Orich	Oak Richness		calculated from field	$405m^2$	ch2
Canopy cover	Densiometer	%	field	$405m^2$	ch2
%upper	% Overstory	%	field	2 samples of $10.2m^2$ in $405m^2$	ch2
%mid	% Midstory	%	field	2 samples of $10.2m^2$ in $405m^2$	ch2
DEN	Number of stems per plot		field	$405m^2$	ch2
SDEN	# of small stems 25 to 75 mm		field	$405m^2$	ch2
MDEN	# of medium stems 75 to 225 mm		field	$405m^2$	ch2
LDEN	# of large stems greater than 225 mm		field	$405m^2$	ch2
BA	Basal area of trees greater than 150 mm	m^2/ha	field	$405m^2$	ch2
PDEN	# of pine stems		field	$405m^2$	ch2
SPDEN	# of small pine stems 25 to 75 mm		field	$405m^2$	ch2
MPDEN	# of medium pine stems 75 to 225 mm		field	$405m^2$	ch2
LPDEN	# of large pine stems > 225 mm		field	$405m^2$	ch2
PBA	Basal area of pine trees > 150 mm	m^2/ha	field	$405m^2$	ch2
HDEN	# of hardwood stems		field	$405m^2$	ch2
SHDEN	# of hardwood stems 25 to 75 mm		field	$405m^2$	ch2
MHDEN	# of hardwood stems 75 to 225 mm		field	$405m^2$	ch2
LHDEN	# of hardwood stems > 225 mm		field	$405m^2$	ch2
HBA	Basal area of hardwood trees > 150 mm	m^2/ha	field	$405m^2$	ch2
HHSDEN	Number of heavy seeding hardwood stems		field	$405m^2$	ch2
HHSBA	Basal area of heavy seeding hardwood trees > 150 mm		field	$405m^2$	ch2
Age	Forest Age	years	field paper records/aerial photography	$405m^2$	ch2
Northness	transformed into a linear north–south gradient, cosine transformation		USGS 10 m DEM	$100m^2$ (down sampled to $1m^2$ for model extrapolation, no adjustment)	ch3

Table A9-16 Continued: Summary of all independent variables used in Dawn Lemke thesis: Alien plants and their invasion of the forested landscape of the southeastern United States.

Code	Variable	Unit	Sources	Resolution	Chapter
Curvature	shape of the landscape, whether it is flat, convex, or concave (increasing positive scores representing increasing concavity)		USGS 10 m DEM	$100m^2$ (down sampled to $1m^2$ for model extrapolation, no adjustment)	ch3
Slope	Slope	degrees	USGS 10 m DEM	$100m^2$ (down sampled to $1m^2$ for model extrapolation, no adjustment)	ch3
Solar	Solar radiation given no cloud cover	Wh/m^2	USGS 10 m DEM	$100m^2$ (down sampled to $1m^2$ for model extrapolation, no adjustment)	ch3
Dist River	Distance to water	m	2009 aerial photo	$1m^2$	ch3
Open100	Proportion of open land within 100 m of plot	proportion	2009 aerial photo	$1m^2$	ch3
Dist Forest	Distance to forest	m	2009 aerial photo	$1m^2$	ch3
Dist Roads	Distance to roads	m	2009 aerial photo	$1m^2$	ch3
Forest06	Proportion of forest within 100 m of plot in 2006	proportion	USGS LULC	$900m^2$ (down sampled to $1m^2$ for model extrapolation, no adjustment)	ch3
Forest00	Proportion of forest within 100 m of plot in 2001	proportion	USGS LULC	$900m^2$ (down sampled to $1m^2$ for model extrapolation, no adjustment)	ch3
Forest90	Proportion of forest within 100 m of plot in 1992	proportion	USGS LULC	$900m^2$ (down sampled to $1m^2$ for model extrapolation, no adjustment)	ch3
NDVI2009	NDVI in 2009		2009 aerial photo	$1m^2$	ch3
NDVI1987	NDVI in 1987		Landsat	$900m^2$ (down sampled to $1m^2$ for model extrapolation, no adjustment)	ch3
NDVI1991	NDVI in 1991		Landsat	$900m^2$ (down sampled to $1m^2$ for model extrapolation, no adjustment)	ch3
NDVI1998	NDVI in 1998		Landsat	$900m^2$ (down sampled to $1m^2$ for model extrapolation, no adjustment)	ch3
NDVI2004	NDVI in 2004		Landsat	$900m^2$ (down sampled to $1m^2$ for model extrapolation, no adjustment)	ch3
NDVI2011	NDVI in 2011		Landsat	$900m^2$ (down sampled to $1m^2$ for model extrapolation, no adjustment)	ch3
DI75	Disturbance Index for 1975		Landsat	$6400m^2$ (down sampled to $900m^2$, based on average)	ch4, ch5
DI90	Disturbance Index for 1990		Landsat	$900m^2$	ch4, ch5

Table A9-16 Continued: Summary of all independent variables used in Dawn Lemke thesis: Alien plants and their invasion of the forested landscape of the southeastern United States.

Code	Variable	Unit	Sources	Resolution	Chapter
DI00	Disturbance Index for 2000		Landsat	900m^2	ch4, ch5
DI90-75	Change in Disturbance Index between 1975 and 1990		Landsat	900m^2	ch4, ch5
DI00-90	Change in Disturbance Index between 1990 and 2000		Landsat	900m^2	ch4, ch5
NDVI75	NDVI in 1975		Landsat	900m^2	ch4, ch5
NDVI90	NDVI in 1990		Landsat	900m^2	ch4, ch5
NDVI00	NDVI in 2000		Landsat	900m^2	ch4, ch5
NDVI90-75	Difference in NDVI between 1975 and 1990		Landsat	900m^2	ch4, ch5
NDVI00-90	Difference in NDVI between 1990 and 2000		Landsat	900m^2	ch4, ch5
CENSUS	Number of people m in 2000	km^{-2}	Census 2000 *TIGER*	Census block (down sample to 900m^2, no adjustment)	ch4, ch5
RD_DIST	Distance to road	m	Census 2000 *TIGER*	900m^2	ch4, ch5
RD_DEN	Density of roads in 2000	km^{-2}	Census 2000 *TIGER*	900m^2	ch4, ch5
MRD_DIST	Distance to major road	m	Census 2000 *TIGER*	900m^2	ch4, ch5
RES ALL	Residential in 2000 or 1990 within a 500 m buffer	binary	USGS LULC	900m^2	ch4, ch5
RES100	Residential presence within a 100 m buffer in 2000	binary	USGS LULC	900m^2	ch4, ch5
RES500	Residential presence within a 500 m buffer in 2000	binary	USGS LULC	900m^2	ch4, ch5
MINT	Minimum temperature from a 30 year average	^{o}C x 10	PRISM	900m^2	ch4, ch5
MAXT	Maximum temperature from a 30 year average	^{o}C x 10	PRISM	900m^2	ch4, ch5
AVET	Average temperature from a 30 year average	^{o}C x 10	PRISM	900m^2	ch4, ch5
RAIN	Average yearly rainfall from a 30 year average	mm	PRISM	900m^2	ch4, ch5
NORTH	Cosine transformation of aspect		USGS National Elevation Dataset	900m^2	ch4, ch5
EAST	Sine transformations of aspect		USGS National Elevation Dataset	900m^2	ch4, ch5
NORTHNESS	Cosine of aspect multiplied by the sine of slope		USGS National Elevation Dataset	900m^2	ch4, ch5
EASTNESS	Sine of aspect multiplied by the sine of slope		USGS National Elevation Dataset	900m^2	ch4, ch5
SLOPE	Slope	degrees	USGS National Elevation Dataset	900m^2	ch4, ch5

Table A9-16 Continued: Summary of all independent variables used in Dawn Lemke thesis: Alien plants and their invasion of the forested landscape of the southeastern United States.

Code	Variable	Unit	Sources	Resolution	Chapter
HILL	Solar radiation given no cloud cover	Wh/m^2	USGS National Elevation Dataset	$900m^2$	ch4, ch5
CURV	shape of the landscape, whether it is flat, convex, or concave		USGS National Elevation Dataset	$900m^2$	ch4, ch5
DEM	Elevation	m	USGS National Elevation Dataset	$900m^2$	ch4, ch5
FC100	Change in forest between 2000 and 1990 within a 100 m buffer	proportion	USGS LULC	$900m^2$	ch4, ch5
FC500	Change in forest between 2000 and 1990 within a 500 m buffer	proportion	USGS LULC	$900m^2$	ch4, ch5
F00 100	Proportion of forest in 2000 with in a 100 m buffer	proportion	USGS LULC	$900m^2$	ch4, ch5
F00 500	Proportion of forest in 2000 with in a 500 m buffer	proportion	USGS LULC	$900m^2$	ch4, ch5
FARM100	Proportion of farming in 2000 with in a 100 m buffer	proportion	USGS LULC	$900m^2$	ch4, ch5
FARM500	Proportion of farming in 2000 with in a 500 m buffer	proportion	USGS LULC	$900m^2$	ch4, ch5
LULC90	Categorical land use in 1990 based on Andersons groupings		USGS LULC	$900m^2$	ch4, ch5
LULC00	Categorical land use in 2000 based on Andersons groupings		USGS LULC	$900m^2$	ch4, ch5
RIV DIS	Distance from a stream	m	USGS	$900m^2$	ch4, ch5
RIV_DEN	Density of streams	km^{-2}	USGS	$900m^2$	ch4, ch5
WATER100	Occurrence of a wetland or stream within 100 m	binary	USGS LULC	$900m^2$	ch4, ch5
WATER500	Occurrence of a wetland or stream within 500 m	binary	USGS LULC	$900m^2$	ch4, ch5
NORTH	Cosine of aspect multiplied by the sine of slope		Piedallu & Gégout, 2008	$1800m^2$	ch6, ch7
SLOPE	Slope	degrees	ESRI, 2009	$1800m^2$	ch6, ch7
CENSUS	Number of people in 2000	km^{-2}	USBOC 2000	Census block (down sample to $1800m^2$, no adjustment)	ch6, ch7
RD	Distance of the FIA site to closest road	m	USBOC 2000	$1800m^2$	ch6, ch7
RIV	Distance to Rivers	m	seamless.usgs.gov	$1800m^2$	ch6, ch7

Table A9-16 Continued: Summary of all independent variables used in Dawn Lemke thesis: Alien plants and their invasion of the forested landscape of the southeastern United States.

Code	Variable	Unit	Sources	Resolution	Chapter
MRD	Distance of the FIA site to closest main road	m	USBOC 2000	1800m^2	ch6, ch7
CITY	Distance to cities in 2000	m	USBOC 2000	1800m^2	ch6, ch7
FOR	Forest in 2001 within a 500 m buffer	proportion	Anderson et al. 1976	1800m^2	ch6, ch7
FARM	Farming in 2001 within a 500 m buffer	proportion	Anderson et al. 1976	1800m^2	ch6, ch7
GRASS	Grassland in 2001 within a 500 m buffer	proportion	Anderson et al. 1976	1800m^2	ch6, ch7
PINE	Pines in 2001 within a 500 m buffer	proportion	Anderson et al. 1976	1800m^2	ch6, ch7
RES	Residential in 2001 within a 500 m buffer	proportion	Anderson et al. 1976	1800m^2	ch6, ch7
FORC	Change in forest (1992 to 2001)	proportion	Anderson et al. 1976	1800m^2	ch6, ch7
PINEC	Change in pines (1992 to 2001)	proportion	Anderson et al. 1976	1800m^2	ch6, ch7
BIO1	Annual Mean Temperature	^oC x 10	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO2	Mean Diurnal Range	Mean of monthly (max temp - min temp) x10	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO3	Isothermality	BIO2/BIO7x100	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO4	Temperature Seasonality	SDx100	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO5	Max Temperature of Warmest Month	°C x 10	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO6	Min Temperature of Coldest Month	°C x 10	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO7	Temperature Annual Range	BIO5-BIO6	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO8	Mean Temperature of Wettest Quarter	°C x 10	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7

Table A9-16 Continued: Summary of all independent variables used in Dawn Lemke thesis: Alien plants and their invasion of the forested landscape of the southeastern United States.

Code	Variable	Unit	Sources	Resolution	Chapter
BIO9	Mean Temperature of Driest Quarter	°C x 10	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO10	Mean Temperature of Warmest Quarter	°C x 10	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO11	Mean Temperature of Coldest Quarter	°C x 10	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO12	Annual Precipitation	mm	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO13	Precipitation of Wettest Month	mm	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO14	Precipitation of Driest Month	mm	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO15	Precipitation Seasonality	CV	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO16	Precipitation of Wettest Quarter	mm	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO17	Precipitation of Driest Quarter	mm	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO18	Precipitation of Warmest Quarter	mm	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
BIO19	Precipitation of Coldest Quarter	mm	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7
ELEV	Elevation	m	Bioclim	1km^2 (down sampled to 1800m^2, using average)	ch6, ch7